

**ADRIANO LOPES ASSUNÇÃO**

**DATA WAREHOUSE: MERCADO DE TRABALHO PARA O GESTOR  
DA INFORMAÇÃO?**

**Trabalho de conclusão de curso  
apresentado à disciplina Pesquisa em  
Informação II, como requisito parcial à  
conclusão do Curso de Gestão da  
Informação, Setor de Ciências Sociais  
Aplicadas, Universidade Federal do  
Paraná.**

**Orientadora: Profa. Ligia Leindorf  
Bartz Kraemer**

**CURITIBA**

**2001**

Um indivíduo sem informações  
não pode assumir responsabilidades,  
mas um indivíduo que recebeu informações  
não pode deixar de assumir responsabilidades.

Jean Carlzon

## SUMÁRIO

<b>LISTA DE FIGURAS</b> .....	iv
<b>LISTA DE SIGLAS</b> .....	v
<b>RESUMO</b> .....	vi
<b>1 INTRODUÇÃO</b> .....	1
<b>2 PROCEDIMENTOS METODOLÓGICOS</b> .....	4
<b>3 SISTEMAS DE APOIO À DECISÃO</b> .....	6
<b>4 O DATA WAREHOUSE</b> .....	12
4.1 CARACTERÍSTICAS DO <i>DATA WAREHOUSE</i> .....	13
4.1.1 Orientação por Assunto.....	13
4.1.2 Integração dos Dados.....	14
4.1.3 Variante Temporal dos Dados.....	14
4.1.4 Não Volatilidade dos Dados.....	15
4.2 ESTRUTURA DO <i>DATA WAREHOUSE</i> .....	16
4.2.1 Granularidade dos Dados.....	17
4.2.2 Metadados.....	20
4.3 MODELAGEM DOS DADOS.....	21
4.4 POVOAMENTO DO <i>DATA WAREHOUSE</i> .....	23
4.4.1 Qualidade dos Dados.....	24
4.4.2 Extração.....	25
4.4.3 Transformação e Integração.....	26
4.5 EXTRAÇÃO DE INFORMAÇÕES DO <i>DATA WAREHOUSE</i> .....	26
4.5.1 Geradores de Consultas e Relatórios.....	27
4.5.2 Sistemas de Informações Executivas (EIS).....	27
4.5.3 Processamento Analítico <i>On Line</i> - OLAP.....	28
4.5.4 <i>Data Mining</i> .....	30
4.6 VANTAGENS DA UTILIZAÇÃO DO <i>DATA WAREHOUSE</i> .....	31
4.7 EXEMPLO DE UTILIZAÇÃO DO <i>DATA WAREHOUSE</i> .....	32
4.8 PROFISSIONAIS ENVOLVIDOS.....	33
<b>5 O GESTOR DA INFORMAÇÃO</b> .....	35
<b>6 PESQUISA DE CAMPO</b> .....	39
<b>7 CONSIDERAÇÕES FINAIS</b> .....	41
<b>REFERÊNCIAS</b> .....	455
<b>DOCUMENTOS CONSULTADOS</b> .....	47
<b>APÊNDICE – FORMULÁRIO DA ENTREVISTA</b> .....	49
<b>ANEXO – CURRÍCULO PLENO DO CURSO DE GESTÃO DA INFORMAÇÃO – RESOLUÇÃO N. 24/98 - CEPE</b> .....	51

## LISTA DE FIGURAS

FIGURA 1 - DIAGRAMA GENÉRICO DE UM SISTEMA .....	7
FIGURA 2 - DIAGRAMA GENÉRICO DE UM SISTEMA DE INFORMAÇÃO .....	8
FIGURA 3 - COMPARAÇÃO ENTRE O AMBIENTE OPERACIONAL E O <i>DATA WAREHOUSE</i> EM RELAÇÃO À ORIENTAÇÃO DE SUA APLICAÇÃO.....	13
FIGURA 4 - COMPARAÇÃO ENTRE O AMBIENTE OPERACIONAL E O <i>DATA WAREHOUSE</i> EM RELAÇÃO À INTEGRAÇÃO DOS DADOS.....	14
FIGURA 5 - COMPARAÇÃO ENTRE O AMBIENTE OPERACIONAL E O <i>DATA WAREHOUSE</i> EM RELAÇÃO À VARIAÇÃO NO TEMPO.....	15
FIGURA 6 - COMPARAÇÃO ENTRE O AMBIENTE OPERACIONAL E O <i>DATA WAREHOUSE</i> EM RELAÇÃO À NÃO VOLATILIDADE DOS DADOS .....	16
FIGURA 7 - ESTRUTURA DO <i>DATA WAREHOUSE</i> .....	17
FIGURA 8 - NÍVEIS DE DETALHES DO <i>DATA WAREHOUSE</i> .....	19
FIGURA 09 - DIAGRAMA DE UM MODELO DIMENSIONAL.....	22
FIGURA 10 - DIMENSÃO DO PRODUTO NORMALIZADO.....	23
FIGURA 11 - CUBO DE DADOS .....	29

## LISTA DE SIGLAS

ABNT	- Associação Brasileira de Normas Técnicas
BI	- <i>Business Intelligence</i> (Inteligência Competitiva)
CELEPAR	- Companhia de Informática do Paraná
DASD	- <i>Direct Acces Storage Device</i> (Dispositivo de Armazenamento de Acesso Direto)
DBA	- Data Bank Administrator (Administrador de Banco de Dados)
DETRAN	- Departamento de Trânsito do Paraná
DWA	- <i>Data Warehouse Administrator</i> (Administrador de <i>Data Warehouse</i> )
EIS	- <i>Executive Information System</i> (Sistema de Informação Executiva)
ISO	- <i>International Organization for Standardization</i> (Organização Internacional de Padronização)
KDD	- <i>Knowledge Discovery in Database</i> (Descobridor de Conhecimento em Banco de Dados)
MD	- Modelo Dimensional
MER	- Modelo Entidade-Relacionamento
OLAP	- <i>On Line Analytical Processing</i> (Processamento Analítico <i>On Line</i> )
PC	- <i>Personal Computer</i> (Computador Pessoal)
SAD	- Sistema de Apoio à Decisão
SGBD	- Sistema de Gerenciamento de Banco de Dados
SIG	- Sistema de Informação Gerencial
SQL	- <i>Structured Query Language</i> (Linguagem de Consulta Estruturada)
SUCESU-PR	- Sociedade dos Usuários de Informática e Telecomunicações do Paraná
UFPR	- Universidade Federal do Paraná

## RESUMO

Apresenta estudo teórico-conceitual sobre sistemas de apoio à decisão, *data warehouse*, formação e perfil profissional do Gestor da Informação. Traça a evolução dos sistemas de apoio à decisão até uma de suas ferramentas mais recentes, o *data warehouse*. Conceitua e caracteriza o *data warehouse*, apresenta sua estrutura, modelagem de dados, processos de povoamento e extração de informações citando algumas ferramentas para a extração de dados, vantagens e um exemplo prático de sua utilização, além de algumas funções e profissionais envolvidos em um projeto de *data warehouse*. Caracteriza o Gestor da Informação apresentando seu perfil profissional, habilidades e estrutura acadêmica da Universidade Federal do Paraná para sua formação. Complementado por uma pesquisa de campo, o estudo verificou, por meio de entrevista, a opinião de profissionais envolvidos em projetos de *data warehouse*, sobre as equipes responsáveis pelo planejamento, implantação e acompanhamento deste ambiente e sobre os padrões e a qualidade dos dados no *data warehouse*. Apresenta como resultado do estudo as atividades que podem ser desenvolvidas pelo Gestor da Informação no ambiente de *data warehouse*, de acordo com as necessidades observadas na literatura, na pesquisa de campo e comparadas com a formação e o perfil deste profissional.

# 1 INTRODUÇÃO

O progresso científico e tecnológico e a explosão da informação ocorridos no século XX ocasionaram mudanças nas sociedades, quer seja no comportamento dos cidadãos, quer seja com relação à forma de atuação das organizações, que procuram criar novos produtos e serviços para suprir a necessidade de consumidores cada vez mais exigentes.

Para se manter no mercado, as organizações precisam monitorar sua concorrência e também o seu histórico de atuação (quantidade produzida, perfil de clientes, meses com maior ou menor índice de vendas, entre outros fatores) estabelecendo suas estratégias por meio da utilização de informações endógenas e exógenas à organização.

JAMIL e NEVES (2000, p. 46) afirmam que a “obtenção da informação e, principalmente, da informação qualificada, é o grande desafio dos nossos tempos”.

Nesse contexto, estão os bancos de dados corporativos e todos os demais recursos da tecnologia da informação, utilizados para prover os executivos com informações estratégicas que lhes permitam tomar decisões corretas.

A utilização de ferramentas de inteligência competitiva (ou *Business Intelligence* - BI) é uma nova tendência tecnológica que se propõe a trabalhar sistemas para coleta e análise de informações que possibilitem simulações e aproximações da realidade para dar suporte à tomada de decisão gerencial. Uma destas ferramentas é o *data warehouse*, um grande armazém de dados cuja finalidade é fornecer relatórios históricos para análise e diagnóstico de situações com vistas a ações frente a previsões e tendências de fatos.

Desta forma, a busca na literatura por temas que pudessem contribuir com o melhor entendimento acerca dessas ferramentas de BI levou à escolha de uma delas, o *data warehouse*, como objeto deste estudo.

O interesse pessoal do pesquisador pela área de ferramentas e tecnologias da informação utilizadas nas organizações como suporte ao processamento operacional e

analítico, assim como a proximidade adquirida com a área de Informática (devido à realização de estágio supervisionado em uma empresa pública de Informática), foram fatores relevantes para a determinação do objeto de estudo.

Em bancos de dados, mais especificamente neste caso, em armazéns de dados ou *data warehouses*, a matéria prima produzida é a informação para a gestão de negócios, utilizando-se de instrumentos e equipamentos tecnológicos. A área de *data warehouse* é, portanto, uma área de atuação que exige diferentes perfis profissionais, cada qual com sua respectiva função mas atuando de forma interdisciplinar, com o intuito de alcançar os objetivos traçados pela organização.

A problemática relacionada à expansão do mercado de trabalho do Gestor da Informação frente às constantes inovações tecnológicas levou à investigação sobre a possibilidade de atuação deste profissional na aplicação da ferramenta de BI *data warehouse*. Estabeleceu-se, portanto, como objetivo geral desta pesquisa: identificar o espaço de atuação do Gestor da Informação em bancos de dados do tipo *data warehouse*.

Para que este objetivo fosse alcançado, fez-se necessária a determinação dos seguintes objetivos específicos:

- a) realizar um estudo teórico-conceitual sobre sistemas de apoio à decisão, *data warehouse* e sobre a formação e o perfil do Gestor da Informação;
- b) identificar atividades que são próprias do Gestor da Informação e que poderiam ser de sua responsabilidade neste tipo de aplicação.

Assim, este relatório é composto por esta Introdução, seção 1, que traça um plano geral da composição da pesquisa, seguido da seção Procedimentos Metodológicos, seção 2, onde a pesquisa é caracterizada e o roteiro metodológico utilizado é apresentado.

A revisão da literatura pertinente, por meio da análise do conteúdo e confrontação das idéias e conceitos encontrados nas diferentes fontes, compôs o referencial teórico do trabalho, distribuído entre as seções 3 e 4.

A seção 3, Sistemas de Apoio à Decisão, apresenta os conceitos básicos dos



elementos que compõem um sistema de informação, bem como a evolução desses sistemas e sua caracterização em relação à tomada de decisão.

Na seção 4, *Data Warehouse*, este tipo de banco de dados é apresentado com suas características, estrutura, povoamento, extração de informações e algumas ferramentas utilizadas neste processo e vantagens da sua utilização, encerrando-se com um exemplo prático de aplicação.

A seção 5, O Gestor da Informação, apresenta seu perfil profissional e a estrutura acadêmica da Universidade Federal do Paraná (UFPR) para a sua formação.

Para suprir a lacuna da literatura pertinente, a seção 6, Pesquisa de Campo, apresenta a opinião de profissionais envolvidos com projetos de *data warehouse*, no que se refere à definição dos profissionais envolvidos em aplicações desse tipo e sua base de trabalho para a estruturação e a qualidade dos dados.

Uma seção dedicada a sintetizar os aspectos mais relevantes da pesquisa é denominada Considerações Finais, seção 7, na qual são apresentadas as possibilidades de atuação do Gestor da Informação no ambiente de *data warehouse*.

A pesquisa encerra-se com uma seção de Recomendações, seção 8, na qual são feitas algumas sugestões ao Curso de Gestão da Informação, e ao profissional Gestor da Informação, bem como são apresentados temas para a continuidade de estudos sobre o tema.

A seção Referências, apresenta os documentos que foram consultados e citados durante a composição do trabalho e a seção Documentos Consultados apresenta os documentos que foram utilizados para fornecer subsídios à construção das idéias, porém não citados.

O formulário de entrevista utilizado na pesquisa de campo compõe o Apêndice e o currículo pleno do Curso de Gestão da Informação é anexado à pesquisa.

## 2 PROCEDIMENTOS METODOLÓGICOS

Esta pesquisa caracteriza-se, segundo os objetivos, como uma pesquisa exploratória-conceitual, pois visa criar maior proximidade de autor e leitor com o tema e, segundo os procedimentos de coleta de dados e as fontes de informação, uma pesquisa bibliográfica em virtude da utilização de literatura para compor o referencial teórico-conceitual do tema.

O levantamento de fontes publicadas em meio tradicional ou eletrônico demonstrou que a quantidade de artigos e obras disponibilizadas na íntegra em meio eletrônico (Internet e bases de dados comerciais) predomina sobre as obras publicadas na forma tradicional (livros) e devido à disponibilização eletrônica da informação, houve maior facilidade no acesso aos documentos.

Para a obtenção de dados não contemplados na literatura e que se mostravam necessários para que os objetivos da pesquisa fossem alcançados utilizou-se da entrevista como instrumento para uma pesquisa de campo.

GIL (1999, p. 177) explica que a entrevista como técnica de coleta de dados é “bastante adequada para a obtenção de informações acerca do que as pessoas sabem, crêem, esperam, sentem ou desejam, pretendem fazer, fazem ou fizeram, bem como acerca das suas explicações ou razões a respeito das coisas precedentes”.

A entrevista foi dividida em três blocos. O primeiro voltado à identificação do entrevistado, sua qualificação e organização a que está vinculado, o segundo bloco, voltado às questões específicas relativas aos profissionais envolvidos com projetos de *data warehouse* (composição da equipe, funções, conhecimentos, dificuldades, competências e habilidades) e o terceiro voltado aos padrões adotados no *data warehouse* e à qualidade de seus dados.

A seleção dos entrevistados (profissionais envolvidos com projetos de *data warehouse*) foi determinada aleatoriamente a partir de uma lista de pessoas participantes de eventos e grupos de estudo, fornecida pelo coordenador do grupo de estudos sobre *Structured Query Language* (Linguagem de Consulta Estruturada -

SQL), da Sociedade de Usuários de Informática e Telecomunicações do Paraná (SUCESU-PR). Os contatos foram efetuados na medida em que as pessoas concordaram em contribuir com a pesquisa, totalizando três profissionais.

Visando garantir o anonimato dos entrevistados, optou-se por não citar seus nomes tanto na análise dos dados quanto na seção de Referências.

A análise das entrevistas foi feita pergunta a pergunta, comparando-se as respostas entre si e relacionando as questões abordadas com a literatura e com o perfil do Gestor da Informação.

Desta análise resultaram as considerações finais e as recomendações.

### 3 SISTEMAS DE APOIO À DECISÃO

Para compreender o que são sistemas de informação, mais especificamente aqueles voltados ao auxílio na tomada de decisão, é preciso observar, primeiramente, alguns conceitos básicos tais como dado, informação e sistema.

São inúmeros os autores que apresentam definições para os respectivos termos, as quais apesar de diferentes em sua construção, levam a um entendimento comum acerca de seu significado; isto levou à seleção de poucos autores, mas que conseguem transmitir um conceito genérico e abrangente sobre os termos que são essenciais para a compreensão da ferramenta *data warehouse*, conforme pode-se observar no quadro 1, em que são apresentados os conceitos de dado e informação, segundo LE COADIC (1996, p. 5-6) e REZENDE e ABREU (200, p. 60).

QUADRO 1 – DEFINIÇÃO DE DADO E INFORMAÇÃO

Termo/Autores	LE COADIC	REZENDE e ABREU
Dado	“em informática chama-se dado a representação convencional, codificada, de uma informação sob uma forma que permite seu processamento eletrônico”	“um elemento de informação, um conjunto de letras, números ou dígitos, que tomado isoladamente não transmite nenhum conhecimento, ou seja, não contém um significado claro”
Informação	“um significado transmitido por meio de uma mensagem inscrita em um suporte espacial-temporal: impresso, sinal elétrico, onda sonora, etc. Essa inscrição é feita graças a um sistema de signos (a linguagem), signo este que é um elemento da linguagem que associa um significante a um significado: signo alfabético, palavra, sinal de pontuação”	“o dado trabalhado, útil, tratado, com valor significativo atribuído ou agregado a ele e com um sentido natural e lógico para quem usa a informação”

Fonte: Elaboração do autor

Para traçar a relação entre dado e informação, pode-se dizer, de forma genérica, que a informação é um dado com valor agregado, ou seja, o dado tornado útil e que leva à compreensão do conjunto ou de algum fato ou acontecimento.

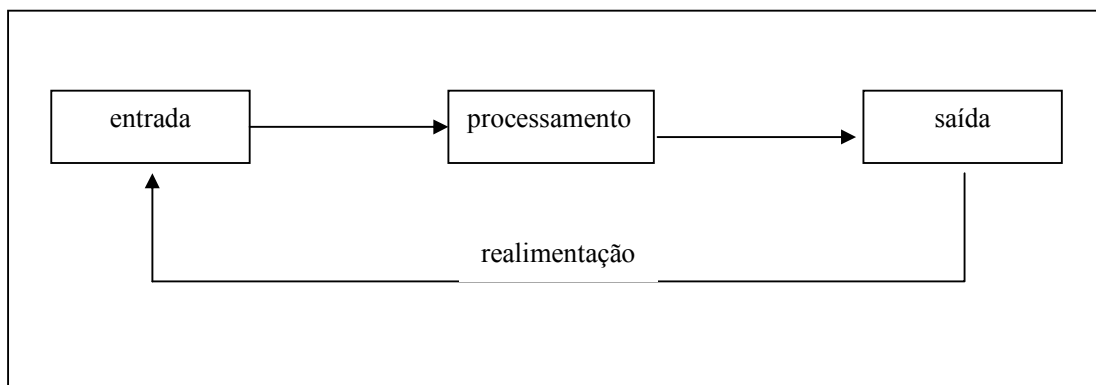
Sistema é definido por OLIVEIRA (2001, p. 23) como “um conjunto de

partes interagentes e interdependentes que, conjuntamente, formam um todo unitário com determinado objetivo e determinada função”.

A representação de um sistema pode ser expressa por um diagrama (Figura 1), onde seus componentes são:

- a) entrada: conjunto de objetos fornecidos ao sistema: material, energia e a informação para a recuperação ou processamento;
- b) processamento: transformação de uma entrada em um produto, serviço ou resultado;
- c) saída: resultado do processamento realizado, na forma de produtos ou serviços;
- d) realimentação ou retroalimentação: reintrodução de elementos julgados pertinentes ao sistema a partir da análise das saídas.

FIGURA 1 - DIAGRAMA GENÉRICO DE UM SISTEMA



Fonte: Elaboração do autor

Em se tratando de sistemas de informação, O'BRIEN<sup>1</sup>, citado por SANTOS (1998, p. 107), os definem como "conjunto de recursos, procedimentos e pessoas que coletam, transformam e disseminam informação em uma organização". Santos, por sua vez, caracteriza sistema de informação como um sistema que acessa dados como

<sup>1</sup> O'BRIEN, James A. **Management information systems: a managerial and use perspective**. Boston: Irwin, 1990.

recursos de informação na sua saída.

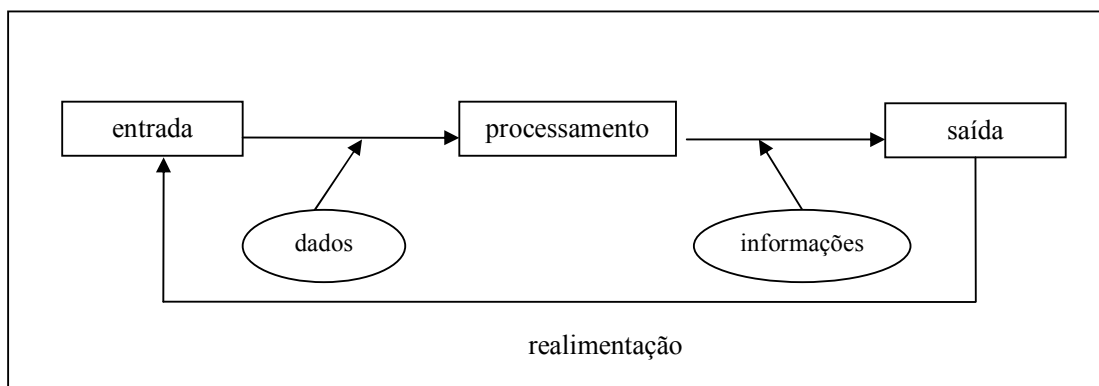
Para NORTON (1996, p. 439), sistema de informação “é um conjunto de regras e procedimentos para o fornecimento de informações precisas e oportunas às pessoas de uma organização, particularmente os gerentes”.

De acordo com CAUTELA e POLONI (1996, p. 23) sistema de informação “é um conjunto de elementos interdependentes (subsistemas), logicamente associados, para que de sua interação sejam geradas informações necessárias à tomada de decisões”.

Dados e informações são os elementos básicos que compõem um sistema de informação, no qual, por sua vez, está embutido o conceito de sistema de um modo geral.

Para destacar a especificidade de um sistema de informação, a figura 2 evidencia a entrada na forma de dados e sua transformação em informação, gerando como saída produtos e serviços de informação.

FIGURA 2 - DIAGRAMA GENÉRICO DE UM SISTEMA DE INFORMAÇÃO



Fonte: Elaboração do autor

Os sistemas de informação podem ser aplicados e classificados de diferentes maneiras no âmbito empresarial, no entanto, neste trabalho é dado ênfase apenas aos sistemas de apoio à decisão.

Os sistemas de apoio à decisão são definidos por POLLONI (2000, p. 31) como “aqueles que tratam de assuntos específicos, estatísticas, projeções, comparações de dados referentes ao desempenho da empresa, estabelecendo parâmetros para novas

ações dentro do negócio da empresa. Estes sistemas são caracterizados pela utilização de pacotes interativos para cálculos e/ou simulações”.

ANDREATTO (1999, p. 1) divide a evolução dos sistemas de apoio à decisão em cinco fases entre o período de 1960 a 1980.

A primeira fase ocorreu no início da década de 1960, quando a computação iniciou suas aplicações voltadas às necessidades individuais, executadas sobre arquivos mestres e caracterizadas por programas e relatórios.

Os arquivos mestres são definidos por INMON (1997, p. 358) como aqueles que “contém o sistema de registro para um determinado conjunto de dados (geralmente destinado a uma aplicação)”.

Ocorrendo por volta de 1965, a segunda fase veio com a explosão do crescimento dos arquivos mestres e das fitas magnéticas, causando problemas como a complexidade de manutenção e desenvolvimento de novos programas, quantidade de *hardware* para manter todos os arquivos mestres e a necessidade de sincronização dos dados a serem atualizados.

A terceira fase veio em 1970 com o surgimento da tecnologia Dispositivo de Armazenamento de Acesso Direto (DASD), em inglês *Direct Access Storage Device*, que substituiu as fitas magnéticas pelo armazenamento em disco. Surge ainda o Sistema de Gerenciamento de Bancos de Dados (SGBD) com o objetivo de tornar o armazenamento e o acesso no DASD mais fáceis para o programador.

A quarta fase ocorreu por volta de 1975 e foi marcada pelo surgimento do processamento de transações *on line*, o que permitiu que o computador fosse utilizado para tarefas antes inviáveis como controlar sistemas de reservas, sistemas de caixas bancários e sistemas de controle de produção entre outros.

As linguagens de quarta geração são aquelas projetadas para permitir que os usuários finais tenham acesso irrestrito aos dados (INMON, 1997, p. 366).

A quinta fase ocorreu em 1980 com as novas tecnologias que começaram a aparecer (PCs e linguagens de quarta geração), por meio das quais o usuário pode então controlar diretamente os sistemas e os dados.

Segundo INMON (1997, p. 5) foi nesta última fase que surgiram os Sistemas

de Informações Gerenciais (SIG) atualmente conhecidos como Sistemas de Apoio à Decisão (SADs) e que consistem no processamento direcionado às decisões gerenciais.

Para CANONGIA *et al.* (2001, p. 8) “nos anos 80 e 90 o destaque foi dado aos sistemas especialistas, que utilizavam inteligência artificial além de apontar soluções, oferecer deduções e pareceres”. Ainda segundo estes autores, no final dos anos 90 o *data warehouse* desponta como “sistema capaz de apoiar a tomada de decisão sob todos os aspectos e em níveis hierárquicos diferenciados, por meio de geração de base de dados integrada, orientada a assuntos de interesse da organização como um todo, estabelecendo base de conhecimento que permite diferentes associações e derivações”.

Para definir e compreender o que é decisão, recorreu-se a PEREIRA e FONSECA (1997, p. 182), para quem

...a decisão não pode ser fragmentada nem polarizada. Ela envolve o ser humano total, nas suas funções lógicas, biológicas e psicológicas. Ela envolve a ética (valor), a estética (sensibilidade), a política (sociedade) e a fé. Não existe decisão essencialmente racional ou emocional. A decisão é sistêmica, é multifacetada, e multidisciplinar.

Conforme OLIVEIRA (2001, p. 38) “a tomada de decisão refere-se à conversão das informações em ação. Portanto, decisão é uma ação tomada com base na análise de informações”.

Os principais objetivos dos sistemas de informações voltados à tomada de decisão, segundo CARVALHO (2001, p. 4) são:

- a) apoio à gerência nas decisões, nos níveis estratégico e tático, por meio de informação resultantes da observação e análise do ambiente tecnológico externo e da avaliação dos impactos das tendências e sinais de mudanças nas áreas de negócio da empresa;
- b) apoiar a revisão das estratégias tecnológicas e empresariais;
- c) ampliar e aprofundar o conhecimento sobre as áreas tecnológicas ligadas aos negócios da empresa;
- d) incentivar a postura estratégica e a visão de futuro nos níveis gerencial e técnico.



Atualmente, devido ao avanço das tecnologias da informação, os gerentes podem utilizar diversas ferramentas para extrair informações sobre seus negócios e tomar decisões baseados no histórico de atuação da empresa. O *data warehouse* é uma destas ferramentas e será o tema do próximo capítulo.

#### 4 O DATA WAREHOUSE

Com a evolução da tecnologia da informação, os sistemas empresariais informatizados foram divididos em:

- a) sistemas operacionais de organizações, ou sistemas que tratam o negócio, ou ainda, sistemas transacionais, os quais dão suporte ao dia a dia na empresa;
- b) sistemas que analisam o negócio, compreendendo os sistemas de apoio à decisão os quais auxiliam os gerentes a interpretar as transações realizadas e traçar estratégias futuras da empresa.

Como os sistemas operacionais estão preparados apenas para as operações quotidianas das empresas e não possuem grande capacidade de armazenamento de dados históricos, os gerentes se vêem frente ao problema da falta de resposta a questões mais complexas de análise da informação. Como solução a estas necessidades, surgiram os programas extratores, que selecionam informações dos sistemas operacionais para utilizá-las em outros ambientes. Conforme apresentado em HISTÓRICO (2001, p. 2), “muitas vezes essas extrações ocorriam em arquivos intermediários, onde as informações sofriam novos tratamentos”. Este processo provocava falhas na integridade das informações, acarretando: falta de credibilidade dos dados, informação publicada com valores diferentes e queda da produtividade.

De acordo com INMON (1997, p. 37) e YAMADA e GIOVANELI (1997, p. 6), dentre outros autores, o *data warehouse* surgiu dessa necessidade de se criar um banco de dados que contivesse o histórico das transações realizadas em um maior período de tempo, pois nos sistemas operacionais as informações limitavam-se a um período de 60 a 90 dias de atuação da empresa.

Desta forma, INMON (1997, p. 33) define *data warehouse* como um “conjunto de dados baseado em assuntos, integrado, não volátil, e variável em relação ao tempo de apoio às decisões gerenciais”.

Para BISPO (1998, p. 1) o *data warehouse* é “um grande banco de dados,

elaborado com a finalidade de dar suporte ao processo decisório, onde os dados que o povoarão são obtidos através dos bancos de dados dos aplicativos operacionais da empresa”.

REZENDE e ABREU (2000, p. 211) define *data warehouse* como “um grande banco de dados que armazena dados de diversas fontes para futura geração de informações integradas, com base nos dados do funcionamento das funções empresariais operacionais de uma organização inteira”.

#### 4.1 CARACTERÍSTICAS DO *DATA WAREHOUSE*

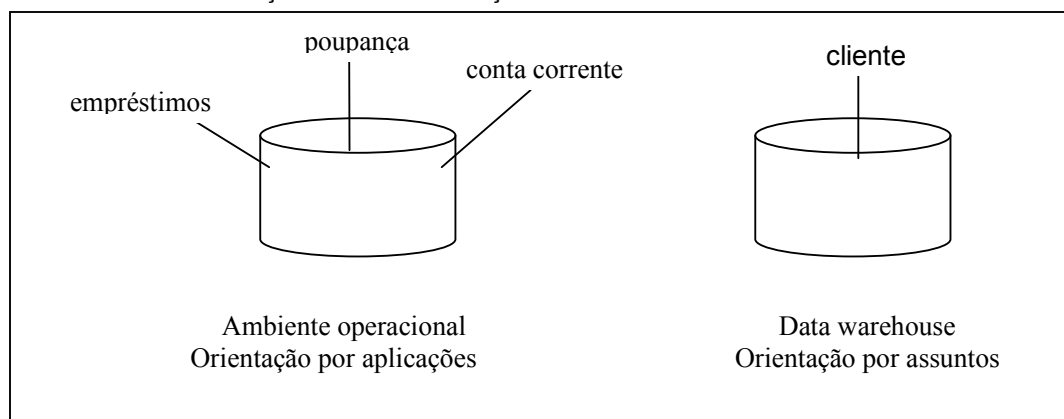
As principais características observadas em um *data warehouse*, em comparação a um banco de dados operacional, são a orientação por assunto, a integração dos dados, variante temporal dos dados e a não volatilidade dos dados.

A seguir estas características serão detalhadas e ilustradas.

##### 4.1.1 Orientação por Assunto

Também denominado orientação por tema, refere-se ao fato de informações estarem agrupadas no *data warehouse* sob temas específicos, em torno das aplicações da empresa, o que pode se observar na figura 3.

FIGURA 3 – COMPARAÇÃO ENTRE O AMBIENTE OPERACIONAL E O *DATA WAREHOUSE* EM RELAÇÃO À ORIENTAÇÃO DE SUA APLICAÇÃO

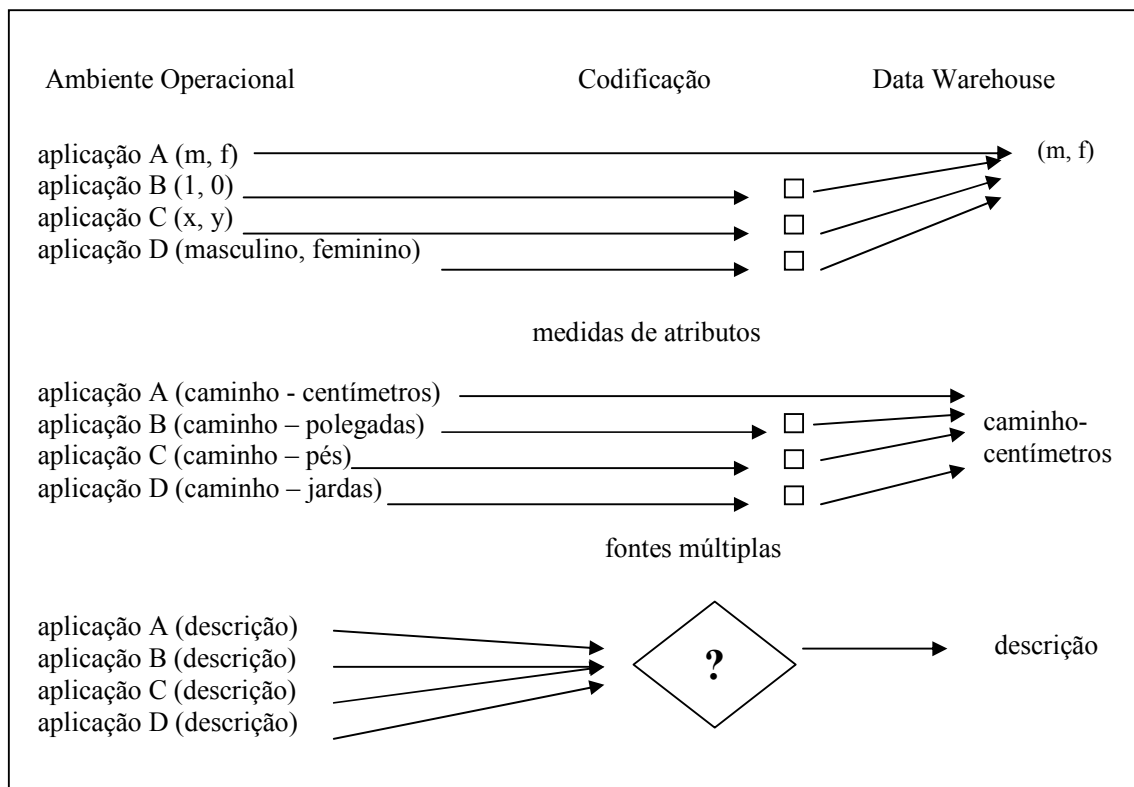


Fonte: YAMADA e GIOVANELI (1997, p.12)

#### 4.1.2 Integração dos Dados

Considerada como a principal característica do *data warehouse*, a integração representa a consistência de atributos (campos) e conteúdos da informação (valores), uma vez que os dados são uniformizados, para então serem carregados, conforme demonstra a figura 4.

FIGURA 4 - COMPARAÇÃO ENTRE O AMBIENTE OPERACIONAL E O *DATA WAREHOUSE* EM RELAÇÃO À INTEGRAÇÃO DOS DADOS



Fonte: Adaptado de INMON (1997, p. 35)

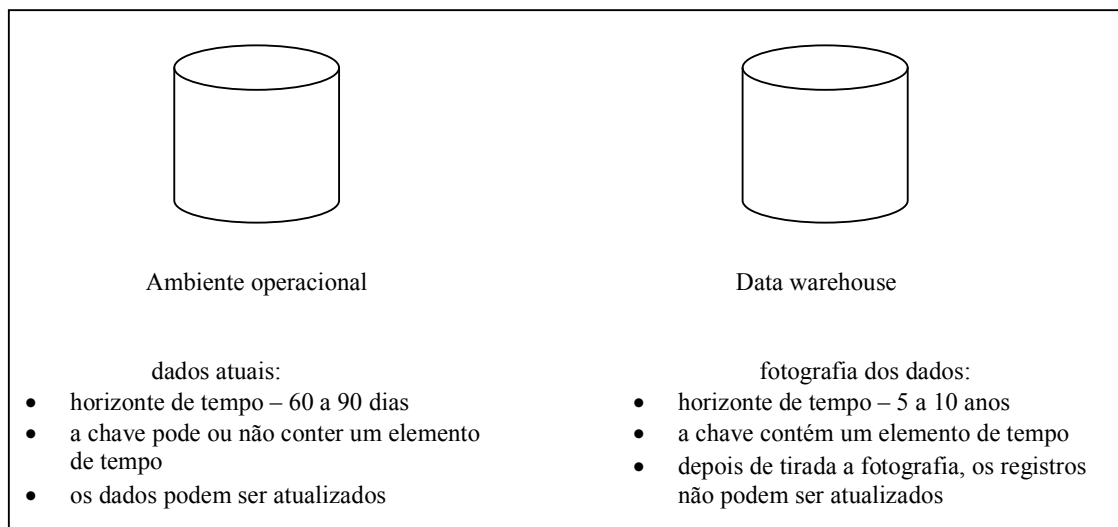
#### 4.1.3 Variante Temporal dos Dados

De acordo com MUSICANTE (2001, p. 7) “ao contrário das aplicações

operacionais, as aplicações informacionais necessitam de dados históricos sobre um período de tempo mais considerável, talvez de cinco a dez anos atrás, derivados das operações que acontecem no dia a dia da organização e que geralmente precisam ser sumarizados e consolidados”.

Como o *data warehouse* é um histórico das transações efetuadas nas organizações, ele contém elementos temporais, ou seja, sempre se refere a um momento específico, conforme ilustra a figura 5.

FIGURA 5 - COMPARAÇÃO ENTRE O AMBIENTE OPERACIONAL E O *DATA WAREHOUSE* EM RELAÇÃO À VARIAÇÃO NO TEMPO



Fonte: INMON (1997, p. 37)

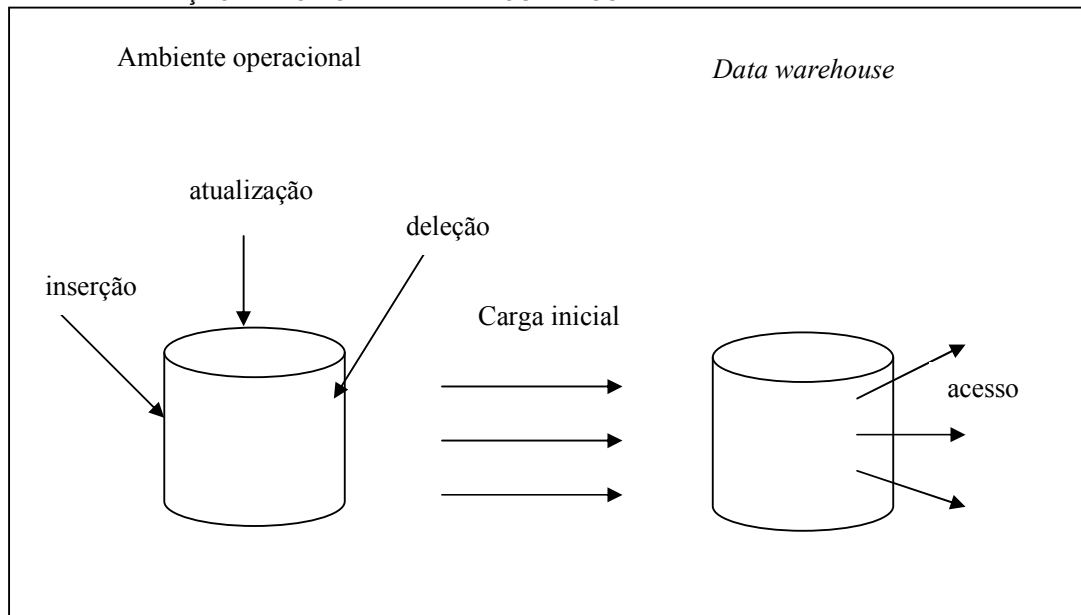
#### 4.1.4 Não Volatilidade dos Dados

No ambiente de um *data warehouse* não é possível fazer modificações nos dados tais como inclusão, alteração ou exclusão, pois após os dados serem carregados, na carga inicial, estes ficam disponíveis apenas para consulta, havendo permissão apenas para a inclusão de mais dados, sem alterar os já existentes.

A carga inicial é a inclusão dos dados no *data warehouse* e o acesso a eles corresponde às consultas para a obtenção de informações, conforme apresentado na

figura 6.

FIGURA 6 - COMPARAÇÃO ENTRE O AMBIENTE OPERACIONAL E O *DATA WAREHOUSE* EM RELAÇÃO À NÃO VOLATILIDADE DOS DADOS



Fonte: YAMADA e GIOVANELI (1997, p. 16)

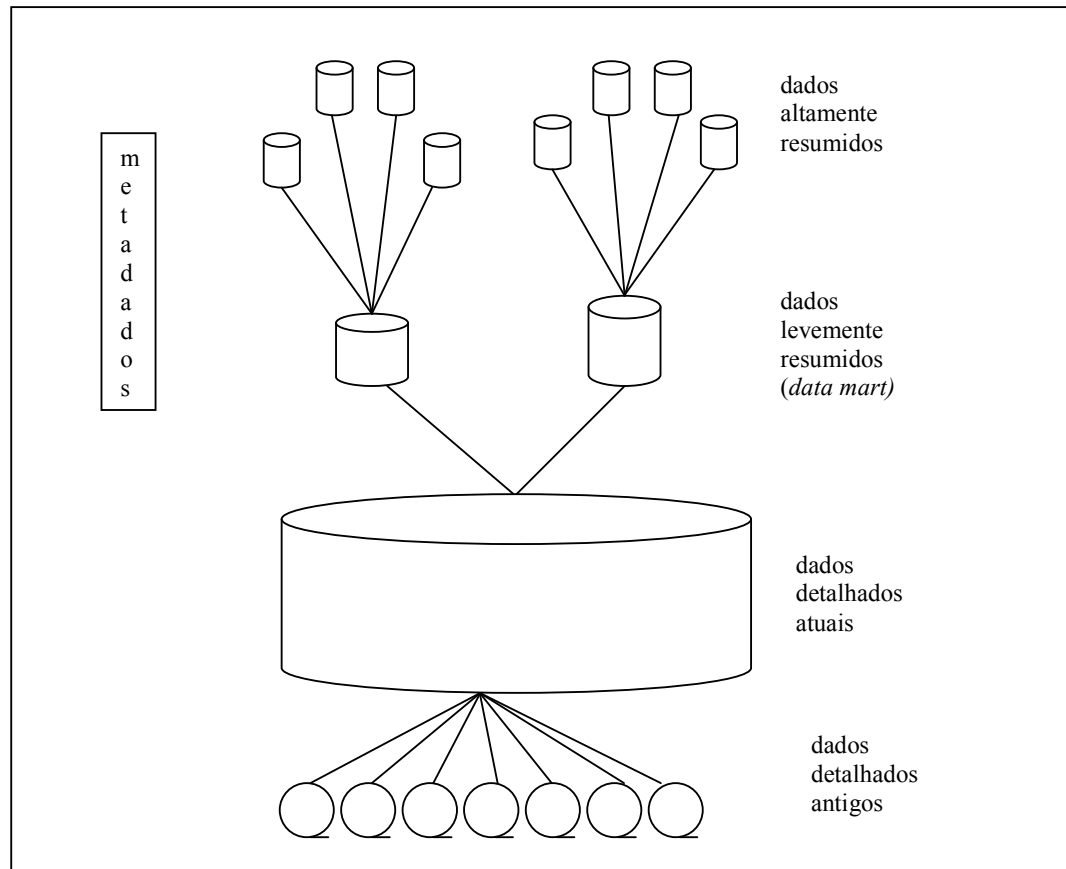
## 4.2 ESTRUTURA DO *DATA WAREHOUSE*

O *data warehouse* possui uma estrutura dividida em diferentes níveis de detalhe e cada nível apresenta características particulares quanto ao detalhamento e utilização (YAMADA e GIOVANELI, 1997, p. 18).

De acordo com INMON e HACKATHORN<sup>2</sup>, citados por YAMADA e GIOVANELI (1997, p. 18), os componentes de um *data warehouse* são: dados detalhados antigos, dados detalhados atuais, dados altamente resumidos, dados levemente resumidos e metadados, conforme mostra a figura 7.

<sup>2</sup> INMON, William H.; HACKATHORN, Richard D. **Como usar o data warehouse**. Rio de Janeiro: Infobook, 1997.

FIGURA 7 - ESTRUTURA DO DATA WAREHOUSE



Fonte: YAMADA e GIOVANELI (1997, p. 18)

Esses componentes podem ser, resumidamente, classificados em duas categorias: granularidade dos dados (compressão ou amplitude de detalhamento) e metadados.

#### 4.2.1 Granularidade dos Dados

A granularidade é o mais importante aspecto de um projeto de *data warehouse*. De acordo com INMON (1997, p. 45) “a granularidade diz respeito ao nível de detalhe ou de resumo contido nas unidades de dados existentes no *data warehouse*”.

O detalhamento dos dados implica diretamente na granularidade, pois quanto mais detalhe, mais baixo o nível de granularidade e quanto menos detalhe, mais alto o nível de granularidade dos dados.

Os dados detalhados atuais são uma das fontes de maior preocupação em um *data warehouse* pois, seu grande volume de dados é armazenado geralmente em discos, uma forma economicamente cara e de difícil gerenciamento, mas por apresentarem acontecimentos recentes que despertam grandes interesses, são muito acessados (YAMADA e GIOVANELI, 1997, p. 19).

Dados detalhados antigos possuem grandes volumes e são pouco acessados, por isso, geralmente, são arquivados em meios de armazenamento de massa, como as fitas magnéticas.

Os dados levemente resumidos não possuem detalhes de baixo nível. Segundo INMON (1997, p. 36) os dados levemente resumidos também podem ser denominados *data mart* ou depósito de dados.

Um *data mart* é um pequeno *data warehouse*, que oferece suporte à decisão a um departamento ou um grupo de pessoas, isto é, como um depósito de dados voltado a uma determinada área do negócio. Ele diferencia-se do *data warehouse* apenas quanto ao tamanho e ao escopo do problema a ser resolvido.

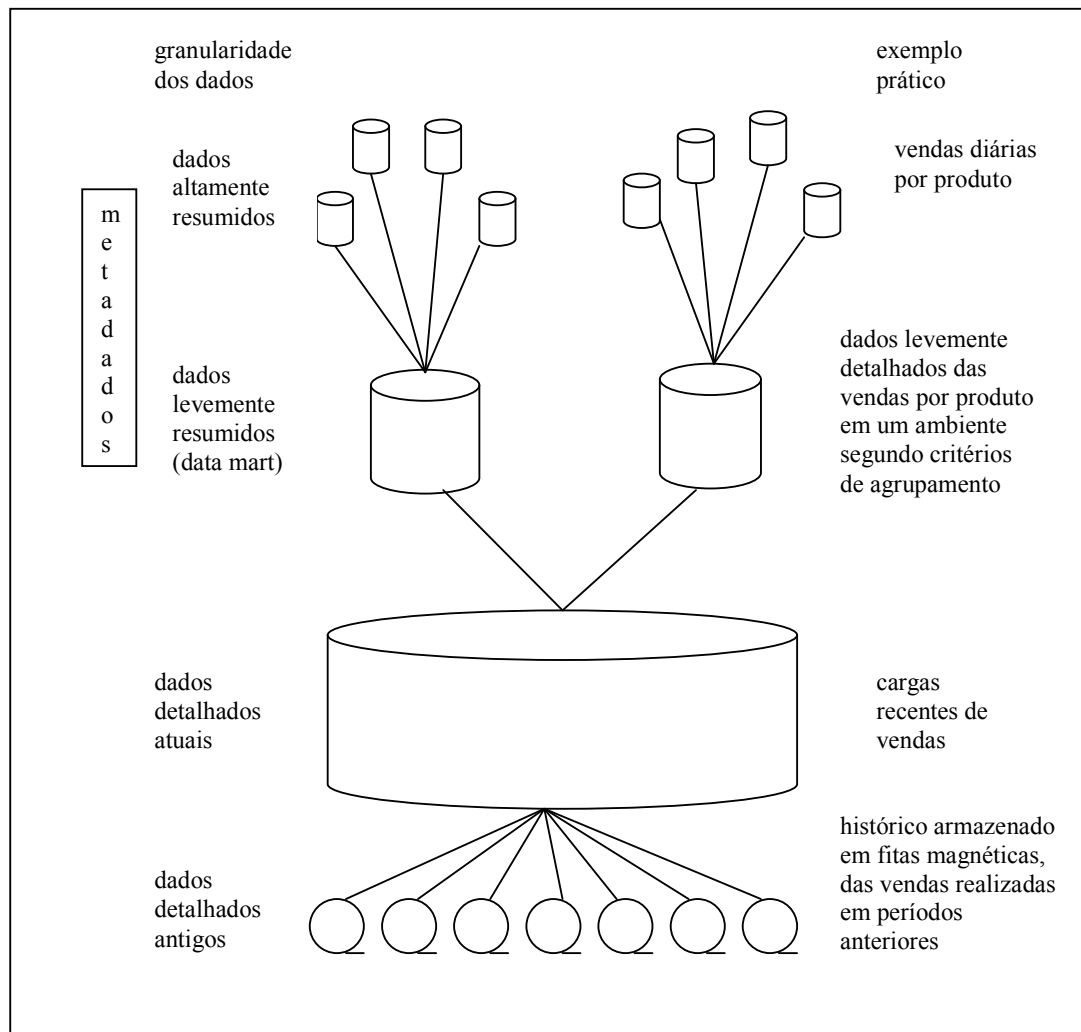
As organizações podem começar a implantação de um *data warehouse* por meio da criação dos *data marts* departamentais, para com o tempo integrá-los e concretizar o *data warehouse*, o que traz como vantagem a redução dos riscos do investimento

Dados altamente resumidos são sucintos e de fácil acesso aos usuários. Segundo YAMADA e GIOVANELI (1997, p. 18), “há a possibilidade de que os dados nesse nível não estejam disponíveis dentro do ambiente do *data warehouse*, apesar disso, continuam pertencendo ao *data warehouse* por mais que a localização física não seja a mesma”.

A figura 8 mostra um exemplo prático de níveis de detalhes que podem ser encontrados no *data warehouse*.



FIGURA 8 – NÍVEIS DE DETALHES DO DATA WAREHOUSE



Fonte: Adaptado de INMON (1997, p. 38)

Quando uma organização possui um grande volume de dados no *warehouse*, faz sentido pensar em dois (ou mais) níveis de granularidade de dados no *data warehouse* (INMON, 1997, p. 49).

A criação de dois ou mais níveis de granularidade no nível detalhado do *data warehouse* permite que vários tipos de consultas sejam atendidos, pois a maior parte do processamento analítico dirige-se aos dados levemente resumidos, que são compactos e de fácil acesso. Para as ocasiões em que um maior nível de detalhe

precisa ser investigado, existe o nível de dados históricos, cujo acesso é caro, incômodo e complexo, mas caso haja necessidade, este nível de detalhe poderá ser alcançado (INMON, 1997, p. 52-53; ANDREATTO, 1999, p. 6).

#### 4.2.2 Metadados

Os metadados constituem-se como uma forma de interação homem-máquina, à medida que são utilizados para orientação do usuário na busca pela informação.

Há na literatura uma unanimidade entre os autores ao afirmar que metadados são dados sobre dados. No entanto, é possível encontrar outras definições tais como a de BERNARDINO (1997, p. 5), para quem o metadado “é uma abstração sobre o dado. É um dado de mais alto nível que descreve um dado de mais baixo nível”.

De acordo com YAMADA e GIOVANELI (1997, p. 21) “os metadados são utilizados para identificar e recuperar os dados solicitados pelos usuários, funcionando como um guia à medida que contém informações sobre quais dados estão disponíveis no *data warehouse*, onde estão localizados e como fazer para acessá-los”.

Para WG SYSTEMS (2002, p. 2), os metadados “são dados a respeito de dados. Exemplos de metadados incluem as descrições de elementos de dados, descrições de tipos de dados, atributos/propriedades, faixas/domínios, métodos e processos”.

É por meio dos metadados que a utilização mais produtiva do *data warehouse* é alcançada.

Segundo INMON (1997, p. 109) os metadados mantêm informações sobre o que está e onde no *data warehouse* por meio dos seguintes fatores:

- a) estrutura dos dados, segundo a visão do programador;
- b) estrutura dos dados, segundo a visão dos analistas de SAD;
- c) fonte de dados que alimenta o *data warehouse*;
- d) transformação sofrida pelos dados no momento de sua migração para o *data warehouse*;
- e) relacionamento entre o modelo de dados e o *data warehouse*;

f) histórico de extrações.

O papel desempenhado pelos metadados no ambiente operacional se difere do seu papel desempenhado no ambiente de *data warehouse*.

No ambiente operacional, os metadados são tratados como algo secundário, relegado ao mesmo nível de importância da documentação.

No ambiente de *data warehouse* os metadados assumem papel de grande destaque, pois são os primeiros elementos que o analista de SAD examina ao planejar o modo de efetuar o processamento informacional/analítico.

#### 4.3 MODELAGEM DOS DADOS

O modelo de dados desempenha papel fundamental para o desenvolvimento interativo no *data warehouse*. Segundo MUSICANTE (2001, p. 57) e ANDREATTO (1999, p. 15) “quando os esforços de desenvolvimento são baseados em um único modelo de dados, sempre que for necessário unir estes esforços, os níveis de sobreposição e desenvolvimento desconexo serão muito baixos, pois todos os componentes do sistema estarão utilizando a mesma estrutura de dados”.

A modelagem de dados consiste na elaboração de um banco de dados usando modelos já consagrados.

De acordo com BISPO (1998, p. 3) “de uma perspectiva lógica, faz-se a modelagem de dados para aumentar a compreensão dos problemas empresariais e identificar os componentes básicos nos quais a solução será construída”.

Dentre os modelos de dados desenvolvidos, pode-se citar o Modelo Entidade-Relacionamento (MER) e o Modelo Dimensional (MD).

No MER os dados estão divididos em tabelas que se relacionam entre si, formando um complexo diagrama. Por este motivo, recomenda-se a utilização do modelo dimensional para a análise dos dados no ambiente gerencial.

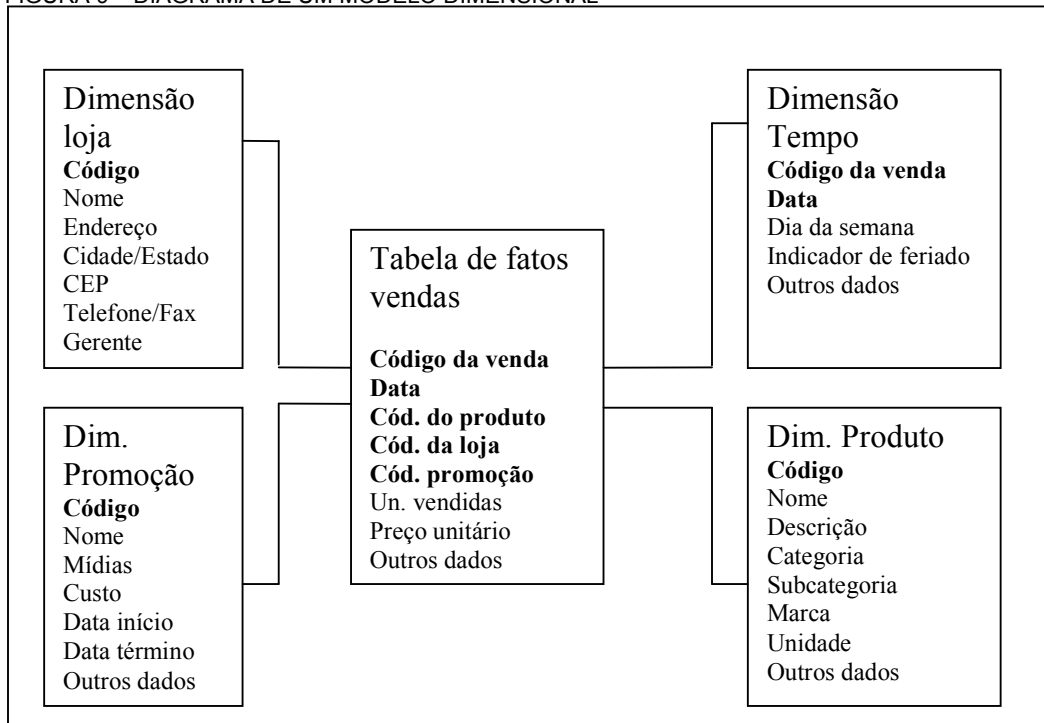
O modelo dimensional também é conhecido como *star schema*, ou esquema estrela, e contém entidades fato, onde estão representados os valores (normalmente numéricos) como objetos do sistema, e entidades dimensão, que circundam a entidade

fato estabelecendo seus relacionamentos.

O modelo dimensional é assimétrico, ou seja, possui uma tabela principal localizada no centro do diagrama e, ao seu redor, outras tabelas menores e que se relacionam entre si. Esta tabela central é a tabela de fatos e as demais tabelas são as tabelas de dimensão.

A representação gráfica deste tipo de esquema é de fácil compreensão e possibilita a visualização dos caminhos de acesso disponíveis para os dados. Este esquema pode ser visualizado na figura 10.

FIGURA 9 – DIAGRAMA DE UM MODELO DIMENSIONAL



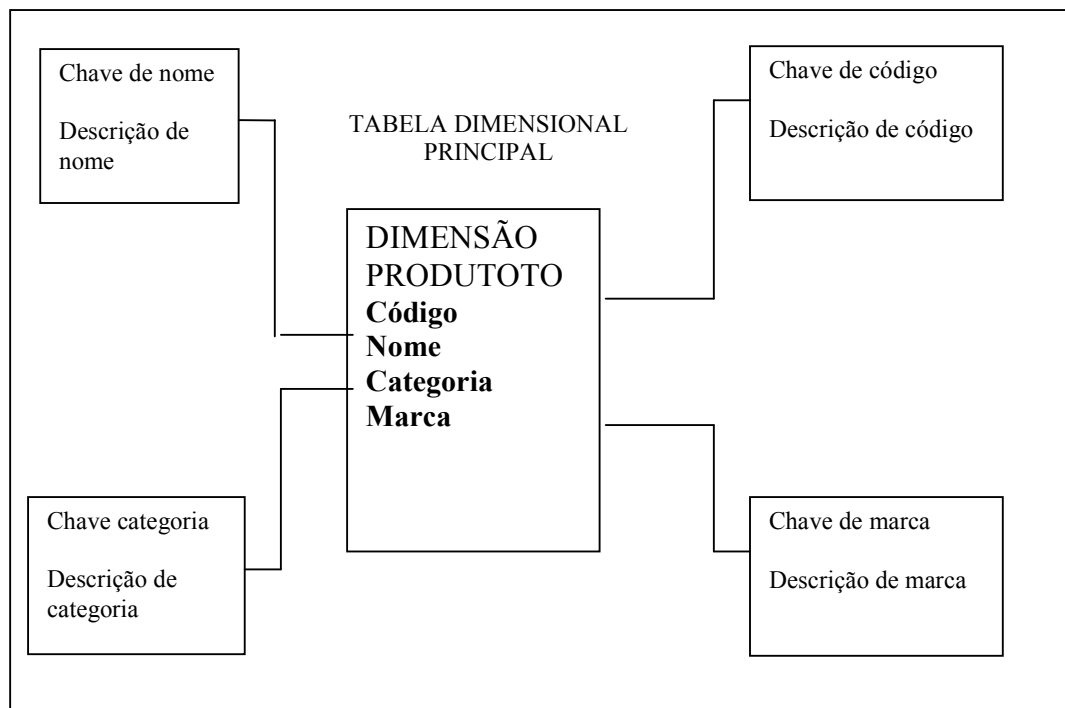
Fonte: BISPO (1998, p. 5)

As dimensões representam as possibilidades de visualização dos dados. As mais comuns nos projetos de *data warehouses* são: tempo (ano, semestre, mês, semana, dia, hora, minuto), geografia (região, estado, cidade, bairro, rua, loja), produto (tipo, marca, sabor) e cliente.

O esquema floco de neve ou *snowflake* é uma extensão do esquema estrela,

em que cada uma das pontas da estrela passa a ser o centro de outras estrelas, pois cada tabela de dimensão é normalizada, quebrando-se a tabela original ao longo das hierarquias existentes em seus atributos (ANDREATTO, 1999, p. 17). Um exemplo de dimensão de produto pode ser visualizado na figura 11.

FIGURA 10 – DIMENSÃO DO PRODUTO NORMALIZADO



Fonte: Elaboração do autor

No exemplo acima, a dimensão produto possui sua hierarquia definida e cada um desses relacionamentos gerou uma nova tabela no esquema floco de neve.

#### 4.4 POVOAMENTO DO *DATA WAREHOUSE*

A extração, limpeza e migração dos dados dos sistemas operacionais existentes para o *data warehouse* são atividades que podem levar ou não ao seu funcionamento efetivo e eficiente (ANDREATTO, 1999, p. 1).

Não há ainda uma ferramenta única que efetue estes processos, mas há diferentes opções de ferramentas especializadas em questões específicas.

De acordo com INMON (1997, p. 73), o *data warehouse* é construído de modo heurístico. “Primeiro o *data warehouse* é povoado com alguns dados, que são examinados minuciosamente pelos analistas de SAD, em seguida, com base no retorno proporcionado pelo usuário final, os dados são modificados e outros dados são adicionados no *data warehouse*”.

#### 4.4.1 Qualidade dos Dados

Como os dados armazenados no *data warehouse* servem para dar suporte aos gerentes na tomada de decisão, estes devem ser totalmente confiáveis, pois diversas são as situações em que a má qualidade destes podem comprometer sua análise, levando a soluções que não seriam as mais adequadas para a resolução de determinado problema (BISPO, 1998, p. 6).

Conforme HUFFORD<sup>3</sup>, citado por BISPO (1998, p. 6), “a qualidade dos dados é o estado da perfeição, validade, consistência e precisão que os dados apresentam durante a sua utilização”.

Os princípios aplicados em programas de gerenciamento pela qualidade total também podem ser utilizados para resolver o problema da qualidade dos dados, pois estes são o resultado do processamento realizado nas organizações, podendo, portanto, ter sua qualidade melhorada.

De acordo com BISPO (1998, p. 7), é preciso realizar um processo de limpeza dos dados. Este processo consiste na extração dos dados do ambiente operacional e a sua transformação até que tenham uma boa qualidade, para então serem carregados no *data warehouse*. Ainda segundo o autor, este processo de limpeza e o custo da sua realização não adicionam valor aos mesmos, no entanto, aumentam

---

<sup>3</sup> HUFFORD, Dane. Data warehouse quality. **DMReview**, Feb./Mar. 1996. Disponível em <<http://www.dmreview.com/master.cfm?NavID=55&EdID=1311#top>> Acesso em: 03 jan. 1998.

sua usabilidade e confiabilidade.

KIMBALL<sup>4</sup>, citado por BISPO (1998, p. 7-9), recomenda o desmembramento da limpeza dos dados em seis passos:

- a) elementarização, ou seja, a divisão dos dados até que estes atinjam a forma mais elementar possível;
- b) padronização dos elementos, como no caso do endereço, onde após este processo, muda-se o “R” para “Rua” e “Sto.” para “Santo”;
- c) verificação da consistência dos elementos padronizados e a verificação de possíveis erros no seu conteúdo;
- d) emparelhamento, que consiste na verificação de duplicidade de entradas para um mesmo registro;
- e) verificação doméstica ou *householding*, que consiste na identificação de pessoas que possuem o mesmo endereço e que possam ter o parentesco comprovado através de seus sobrenomes;
- f) documentação dos resultados dos passos anteriores e o seu armazenamento no banco de dados para metadados.

#### 4.4.2 Extração

De acordo com ANDREATTO (1999, p. 3), há varias possibilidades para extração que permitem balancear o desempenho, restrições de tempo e de armazenamento. Ainda segundo este autor, “os programas que criam os arquivos de extração podem ser incorporados a um ponto apropriado deste esquema de processamento”.

As rotinas de extração devem possibilitar o isolamento de dados que foram incluídos e atualizados desde a última extração, sendo este processo conhecido como

---

<sup>4</sup> KIMBALL, Ralph. **The data warehouse toolkit**: practical techniques for building dimensional Data Warehouses. New York: J. Wiley, 374p.

*refresh*. Conforme DAL'ALBA<sup>5</sup>, citado por ANDREATTO (1999, p. 1), “a melhor política de *refresh* deve ser avaliada pelo administrador do *data warehouse*, que deve levar em conta características como as necessidades dos usuários finais, tráfego na rede e períodos de menor sobrecarga, tanto das origens dos dados quanto do *data warehouse*, deve-se considerar que os períodos de sobrecarga podem variar para cada origem de dados”.

#### 4.4.3 Transformação e Integração

A criação do *data warehouse* não consiste apenas em extrair dados dos sistemas operacionais e inseri-los no *warehouse*.

Estes dados precisam passar por processos de transformações, sendo que o primeiro refere-se à limpeza ou filtragem dos dados, no qual o objetivo é garantir a integridade, deixando-os consistentes antes de serem carregados no *data warehouse*.

O segundo passo é a homogeneização dos dados através da aplicação de metodologias de comparação das representações, nas quais possam ser identificadas as semelhanças e conflitos de modelagem (ANDREATTO, 1999, p. 2).

Pode-se citar como exemplo de limpeza dos dados: correção de erros de digitação, descobertas de violações de integridade, substituição de caracteres desconhecidos e padronização de abreviações e de medidas (ANDREATTO, 1999, p. 2).

Segundo MUSICANTE (2001, p. 28) “para realizarem estas transformações, os programas de integração de dados utilizam informações armazenadas no depósito de metadados”.

#### 4.5 EXTRAÇÃO DE INFORMAÇÕES DO *DATA WAREHOUSE*

Há várias formas de extração de informações do *data warehouse*. De acordo

---

<sup>5</sup>DAL'ALBA, Adriano. **Um estudo sobre Data Warehouse**. Disponível em: <<http://www.geocities.com/siliconvalley/port/5072/>> Acesso em: 01 set. 1999.



com ANDREATTO (1999, p. 35) e MUSICANTE (2001, p. 16), as ferramentas mais comuns são: geradores de consultas e relatórios, Sistemas de Informações Executivas – EIS (do inglês *Executive Information Systems*), ferramentas de Processamento Analítico *On Line* (OLAP) e *data mining*.

#### 4.5.1 Geradores de Consultas e Relatórios

Considerada a primeira geração de ferramentas para acessar os dados, pois permitem a consulta *ad hoc*, possibilitando através de menus e botões, a especificação dos elementos de dados, condições, critérios de agrupamentos, entre outros atributos (MUSICANTE, 2001, p. 16).

A consulta *ad hoc* é um tipo de consulta eventual, ou seja, qualquer consulta que não possa ser determinada antes do momento da consulta ser emitida. Uma consulta que consiste em *Structured Query Language* (SQL) construído dinamicamente, em geral, por ferramentas de consulta residentes na estação de trabalho do usuário final.

Conforme YAMADA e GIOVANELI (1997, p. 39), a própria ferramenta gera uma busca no banco de dados, extrai os dados solicitados e efetua os cálculos adicionais e apresenta os resultados de forma clara. Ainda segundo as autoras, estas ferramentas são mais apropriadas quando são necessárias respostas do tipo “o que aconteceu?” ou “como se comparam as vendas dos produtos X e Y do mês passado com as vendas do mesmo mês de anos anteriores?”.

Esta ferramenta é, normalmente, limitada a médias, totalizações, desvio padrão e outras funções básicas de análise (MUSICANTE, 2001, p. 16; YAMADA e GIOVANELI, 1997, p. 39).

#### 4.5.2 Sistemas de Informações Executivas (EIS)

Esta ferramenta apresenta uma visualização mais simplificada, altamente consolidada e na maioria das vezes, estática.

É destinada especificamente à satisfação das necessidades de executivos de alto nível, eliminando intermediários entre executivos e computadores, permitindo o acesso ao *data warehouse* de forma direta (MUSICANTE, 2001, p. 16).

De acordo com REZENDE e ABREU (2000, p. 201) os EISs “são desenvolvidos de modo que se enquadrem na cultura, filosofia e políticas e no modelo de gestão da empresa. Filtram, resumem e acompanham dados ligados ao controle do desempenho de fatores críticos para o sucesso do negócio”. Ainda conforme estes autores, esta tecnologia tem como prerrogativa a capacidade de *drill-down*, ou seja, recurso de aprofundamento em detalhes, de acordo com as necessidades dos executivos.

Conforme INMON (1997, p. 257) alguns dos usos tradicionais dos EISs são:

- a) análise e investigação de tendências;
- b) mensuração e rastreamento de indicadores de fatores críticos;
- c) análise prospectiva;
- d) monitoramento de problemas;
- e) análise da concorrência.

#### 4.5.3 Processamento Analítico *On Line* - OLAP

O processamento analítico *on line* (OLAP – *On Line Analytical Processing*) permite a extração de forma rápida e interativa, de informações dos dados armazenados no *data warehouse*, independente do tamanho e da complexidade do banco de dados (MUSICANTE, 2001, p. 42).

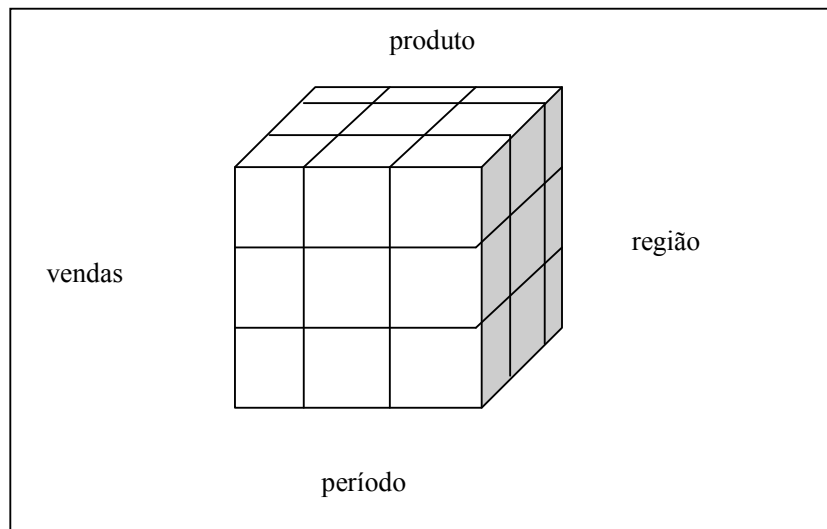
YAMADA e GIOVANELI (1997, p. 40) ressaltam que as primeiras soluções OLAP estavam baseadas em bancos de dados multidimensionais que utilizavam estruturas cúbicas, denominadas hipercubos ou *array* multidimensional, e armazenavam os dados de forma que associações entre eles fossem efetuadas através de múltiplas dimensões.

INMON, TERDEMAN e IMHOFF (2001, p. 242), definem cubo de dados como “um objeto tridimensional composto de colunas e fileiras de células. As margens

do cubo são as dimensões do espaço de análise”.

A figura 11 ilustra um exemplo no qual a visualização do total de vendas a um cliente em determinado período de tempo é um problema dimensional. Quando se acrescenta o dado da localização geográfica é criada a terceira dimensão e a linha de produtos pode ser a quarta.

FIGURA 11 – CUBO DE DADOS



Fonte: YAMADA e GIOVANELI (1997, p. 32)

Esta ferramenta permite que sejam elaboradas consultas *ad hoc*, não restringindo o usuário a consultas pré-definidas.

Com os resultados obtidos na consulta, o usuário pode detalhar cada informação, navegando pelos dados para compreender os motivos dos resultados.

Conforme MUSICANTE (2001, p. 43) o usuário pode utilizar as seguintes operações para refinar sua pesquisa:

- a) *drill-down*: que possibilita ir de um nível genérico para um nível detalhado;
- b) *drill-up*: é o processo inverso do *drill-down*, ou seja, vai do nível detalhado para o nível genérico;
- c) *slice and dice*: permite que os dados sejam visualizados sob diferentes pontos de vista, tal como um cubo a fim de “fatiá-lo” de acordo com as necessidades;

- d) pivoteamento: alteração da forma como os dados podem ser visualizados, como por exemplo, pode se visualizar tanto a quantidade de vendas de cada vendedor dividida por produto, quanto a quantidade de vendas de cada produto dividida por vendedor.

#### 4.5.4 *Data Mining*

As ferramentas de *data mining* (mineração de dados) são especializadas na procura de padrões nos dados. Segundo MUSICANTE (2001, p. 16), “essa busca pode ser efetuada automaticamente pelo sistema ou interativamente com um analista, responsável pela geração de hipóteses (todo o processo é conhecido como *Knowledge Discovery in Database – KDD*)”.

Para ANDREATTO (1999, p. 38) “*data mining* é o processo de extrair informação válida, previamente desconhecida e de máxima abrangência a partir de grandes bases de dados, usando-as para efetuar decisões cruciais”.

A integração de técnicas de *data mining* sobre o ambiente de *data warehouse* permite diversas aplicações, que já vêm sendo utilizadas em diversos segmentos de negócios tais como manufatura, automação de pedidos de remessas, varejo, gerenciamento de inventários, financeiro, análise de risco, transporte, gerenciamento de frotas e telecomunicações, entre outras (ANDREATTO, 1999, p. 38).

Para YAMADA e GIOVANELI (1997, p. 43) o *data mining* “pertence a uma categoria de ferramentas onde, ao invés de se fazerem perguntas, entrega-se para a ferramenta grandes quantidades de dados e se pergunta se existe algo de interessante nos mesmos, como uma tendência ou um agrupamento, por exemplo”.

A forma de atuação do *data mining* pode ser comparada a de um consumidor indo às compras em uma loja de departamentos ou um *shopping center*. Ele não sabe o que adquirir até o momento em que o produto aparece para ser comprado. E o consumidor estava lá para isso.

Outra forma de aplicação desta ferramenta é a elaboração de perfis para produtos e serviços. Imagine-se que uma instituição financeira necessite conhecer as

características de clientes que pagam seus empréstimos em dia. Pode-se descobrir que este cliente tem uma renda anual acima de R\$ 60.000,00, é casado e cliente há mais de dez anos. Tem-se assim um perfil de bons pagadores e a instituição pode diminuir o risco de inadimplência, concedendo crédito apenas aos clientes cujo perfil se encaixa no estabelecido pelo *data mining*.

#### 4.6 VANTAGENS DA UTILIZAÇÃO DO *DATA WAREHOUSE*

Após esta explanação sobre o *data warehouse*, convém ressaltar a sua importância no processo de tomada de decisão, pois além de permitir uma visão global da organização, ele contém os dados sobre o negócio necessários que permitem a análise de informações para a tomada de decisão.

Com esta tecnologia, os gerentes podem executar análises complexas em dados históricos e operacionais sem diminuir o desempenho do banco de dados principal da empresa.

De acordo com STAIR<sup>6</sup>, citado por REZENDE e ABREU (2000, p. 212), a principal vantagem do *data warehouse* “é a possibilidade de relacionar dados de forma nova e criativa”, ressalta, no entanto, que “devem ser observados os impactos desta replicação e a redundância de dados”.

A partir dos dados armazenados no *data warehouse* podem ser obtidas informações que permitem traçar as tendências futuras, de forma mais segura pois estão baseadas no contexto histórico da organização.

Outra vantagem quanto à sua utilização, refere-se à redução do tempo despendido na geração de consultas, pois como os dados estão armazenados em um único local, não há necessidade de se efetuar pesquisas exaustivas em diferentes bancos de dados.

Para YAMADA e GIOVANELI (1997, p. 54) “se por um lado a implantação do *data warehouse* possui um custo muito elevado, por outro ele reduz

---

<sup>6</sup> STAIR, Ralph M. **Princípios de sistemas de informação: uma abordagem gerencial**. 2. ed. Rio de Janeiro: LTC, 1998. 451 p.

significativamente os gastos, se considerarmos a redução do tempo despendido e a recursos utilizados para obtenção da informação”.

Para NIGRO<sup>7</sup>, citado por YAMADA e GIOVANELI (1997, p. 58), “o *data warehouse* é um investimento com resultados garantidos a médio ou longo prazo. Todavia, podemos dizer que a curto prazo os primeiros benefícios visíveis consistem na diminuição da granularidade dos dados e na realização de consultas anteriormente impossíveis de realizar”.

#### 4.7 EXEMPLO DE UTILIZAÇÃO DO *DATA WAREHOUSE*

Para melhor compreensão de como o *data warehouse* pode auxiliar os gerentes na tomada de decisão, será apresentado aqui o caso do *Wal-Mart*, o que é considerado um exemplo de correta utilização das tecnologias da informação mais citados na literatura sobre este tema.

O *Wal-Mart* foi fundado em 1962, nos Estados Unidos e chegou ao Brasil em 1995. Desde sua chegada, a empresa já investiu 500 milhões de dólares. No ano passado, ocupou o sexto lugar no *ranking* nacional do varejo, com uma fatia de 1,8% do mercado. Nos Estados Unidos, em 1999, o *Wal-Mart* teve 30% de um mercado de 379 bilhões de dólares.

Quando um cliente compra uma caixa de lenços descartáveis em um das 21 lojas da rede no Brasil, ao passá-la no leitor de códigos de barras, a informação viajará aproximadamente 13 mil Km, direto para a sede da empresa na cidade de Bentonville, Arkansas, Estados Unidos. De lá o sistema emite aviso para que o centro brasileiro de distribuição reponha a caixa de lenços, e isto acontece em uma fração de segundos, sendo que todo este processo ocorre no bando de dados operacional.

O *data warehouse* do *Wal-Mart* é sete vezes maior que todo o conteúdo da *Library of Congress* (se todos os documentos dessa biblioteca fossem digitalizados,

---

<sup>7</sup> NIGRO, Márcio. O melhor caminho até seu cliente. **Byte Brasil**, São Paulo, Globo, n. 1, p. 44-52, jan. 1997.

daria um total de 20 trilhões de *bytes*) e possui armazenados 140 trilhões de *bytes* em dados.

Por meio da utilização de ferramentas de extração de informações, é possível garimpar neste bando de dados, respostas aos mais diferentes tipos de perguntas tais como: qual o momento do dia em que o consumidor chinês compra mais chá? Que produto deve ser mudado de lugar na gôndola para alcançar os olhos do consumidor coreano? Pode-se inclusive descobrir que, ao contrário do que ocorre em outros países, no Brasil os consumidores têm o hábito de ir às compras acompanhados da família.

Outro exemplo de como o *Wal-Mart* tem utilizado eficientemente o *data warehouse* é citado em YAMADA e GIOVANELLI (1997, p. 56), que foi a constatação de que uma faixa de consumidores entre 25 e 30 anos que compravam fraldas às sextas-feiras, também compravam cervejas. A possível justificativa para este fato é que muitos desses consumidores são homens casados, que ao saírem do trabalho, passam no supermercado para comprar fraldas para os filhos e aproveitam para comprar cervejas para o final de semana. Com estas informações, a loja colocou os dois produtos um ao lado do outro e realizou promoção de cerveja nas sextas-feiras. O resultado foi positivo e empresa obteve um aumento de 30% nas vendas da bebida.

#### 4.8 PROFISSIONAIS ENVOLVIDOS

MIORELLI (2000, p. 3) propõe a definição de um grupo gestor da informação que participe durante todo o desenvolvimento do projeto. Segundo o autor este grupo “pode ser considerado como os usuários que dominam a área de negócio e que tomam decisões, juntamente com o pessoal de informática”.

Para ANDREATTO (1999, p. 45) “os processos de extração, filtragem, carga e recuperação dos dados são bastante complexos, exigindo que pessoas altamente capacitadas façam parte do projeto para que os objetivos sejam atingidos no menor espaço de tempo possível e sem gastos de recursos desnecessários”.

Encontra-se em INMON (1997) e INMON; TERDEMAN e IMHOFF (2001) as seguintes funções e profissionais envolvidos com o planejamento, implantação e

acompanhamento do *data warehouse* e suas respectivas atribuições:

- a) administrador de bancos de dados (DBA) – é o profissional encarregado de cuidar e monitorar diariamente os bancos de dados;
- b) administrador de dados – é o responsável pela especificação, aquisição e manutenção dos *softwares* de gerenciamento de dados e pelo projeto, validação e segurança dos arquivos ou bancos de dados;
- c) administrador de *data warehouse* (DWA) – é a função organizacional designada para criar e manter o *data warehouse*;
- d) analista de SAD - é o usuário do *data warehouse*, que é antes de mais nada, uma pessoa de negócios, e além disso, um técnico cuja principal tarefa consiste na definição e descoberta de informações usadas no processo corporativo de tomada de decisões;
- e) explorador – é o usuário final de SAD que opera com base randômica procurando padrões, associações e outros relacionamentos não observados anteriormente em grandes quantidades de dados detalhados;
- f) fazendeiro – é um usuário de SAD que seguidamente observa pequenas quantidades de dados e freqüentemente encontra o que está procurando;
- g) minerador de dados – é a pessoa que está engajada na mineração de dados.

Observa-se na literatura que há vários perfis profissionais envolvidos em um projeto de *data warehouse*, cada qual com sua respectiva denominação e atribuições, porém sem nenhuma menção ao Gestor da Informação especificamente.

Desta forma, a próxima seção, O Gestor da Informação, traçara o perfil deste profissional e a estrutura acadêmica oferecida pela Universidade Federal do Paraná, para que na seção Considerações Finais sejam apresentadas as possíveis atribuições para este profissional no ambiente de *data warehouse*.



## 5 O GESTOR DA INFORMAÇÃO

Conforme apresentado em GUIA (1999, p. 1) a criação do curso de Gestão da Informação, em 1998 pela Universidade Federal do Paraná (UFPR) “resultou da percepção da necessidade de se formarem profissionais capazes de solucionar, com eficiência, problemas de informação existentes nos diversos setores da sociedade”. Quando de sua criação, esta formação era inédita no Brasil e estava “engajada na dinâmica das transformações mundiais e no conceito e potencialidades de trabalho atuais que a informação assume na sociedade globalizada”.

No entanto, atualmente já podem ser encontrados no Brasil, cursos voltados à Gestão da Informação, porém como ênfase nos cursos de Administração, Biblioteconomia e Sistemas de Informação. No âmbito internacional também se pode cursar na graduação o curso de Gestão da Informação ou *Information Management*, que é lecionado na *Queen Margareth University College*<sup>8</sup> localizada na cidade de Edimburgo, na Escócia.

De acordo com o apresentado em GUIA (1998, p. 2) o Gestor da Informação “tem, em sua formação a oportunidade de direcionar-se pessoal e profissionalmente para:

- a) atender às demandas de trabalho com a geração, a análise, o controle, a transmissão, a distribuição e a utilização da informação;
- b) utilizar as tecnologias a favor do uso efetivo e qualitativo da informação;
- c) atuar em unidades de informação de qualquer natureza;
- d) oferecer serviços e produtos de informação de forma autônoma;
- e) atuar criticamente junto aos contextos que interferem no fluxo, custo e uso da informação.

Ainda segundo este guia, o curso prevê o desenvolvimento de habilidades do Gestor da Informação nos níveis profissional e pessoal, conforme apresentados a seguir.

---

<sup>8</sup> [www.qmuc.ac](http://www.qmuc.ac).

Profissionalmente, o gestor da informação deve ser capaz de:

- a) reconhecer e aplicar teorias e paradigmas da informação;
- b) identificar e explorar fontes de informação;
- c) avaliar a qualidade das fontes de informação, sob os parâmetros de exatidão, atualidade, abrangência, formatos disponíveis e orientação às necessidades do cliente;
- d) adicionar valor ao processo de coleta de informações;
- e) organizar e sistematizar a informação útil a cada cliente;
- f) utilizar processos de análise, interpretação e representação da informação;
- g) coletar e conectar informações dispersas de modo a originar novas informações e conhecimentos;
- h) utilizar a tecnologia como vetor para conectar pessoas, organizações, documentos e informações;
- i) diagnosticar problemas de informação do cliente, definindo quando, como e, mesmo, se a informação deve ser armazenada.
- j) dominar habilidades para:
  - acessar e adquirir informações em qualquer suporte e formato;
  - “navegar” nas redes tradicionais e eletrônicas disponíveis;
  - intercambiar informações entre sistemas de informação existentes;
  - identificar pessoas e organizações como fontes de informação;
  - identificar, localizar e analisar dados não cobertos por sistemas tradicionais de informações.

Pessoalmente, o gestor da informação deve ser capaz de:

- a) inserir-se e contextualizar sua atuação no ambiente no qual trabalha;
- b) coordenar atividades em equipes multidisciplinares;
- c) comunicar-se de forma efetiva;
- d) negociar e vender produtos e serviços de informação;
- e) promover liderança e visão;
- f) orientar-se para clientes;

- g) assumir papel de consultor para problemas de informação;
- h) encorajar e instrumentalizar os indivíduos a identificar e utilizar recursos informacionais.

A utilização da informação como recurso estratégico, requer a atuação de um profissional voltado para o estudo dos fluxos da informação e a promoção de seu uso de forma custo-efetiva.

Sob o ponto de vista da segurança da informação FONTES (2001, p. 2) define Gestor da Informação como “a pessoa responsável pela liberação (ou não) da informação para toda a empresa e também deve ser o responsável pela continuidade do negócio no que depende daquela informação”. Ainda segundo o autor, o Gestor da Informação consciente de sua função, é um aliado para a obtenção e manutenção de recursos para a proteção da informação e deve estar definido na política de segurança da informação da empresa.

Com uma abordagem mais ampla, é apresentado em POLÍTICA (2001, p. 2) que o Gestor da Informação é “o indivíduo responsável para fazer decisões em nome da organização no que diz respeito ao uso, à identificação, e à proteção de um recurso específico da informação”.

O Gestor da Informação tem uma formação voltada para a atuação em equipes multidisciplinares, podendo atuar nos processos de geração e uso de recursos de informação, podendo diagnosticar, propor e implementar ações para a solução de problemas de informação (GUIA, 1998, p. 5).

A formação multidisciplinar do Gestor da Informação fica evidenciada quando se analisa o *roll* das 71 disciplinas que compõem o currículo do curso e que, dividido em cinco linhas, apresenta-se da seguinte forma:

- a) 33,8% das disciplinas são voltadas para o tratamento da informação (24 disciplinas);
- b) 29,6% das disciplinas são voltadas para a área de Administração (21 disciplinas);
- c) 23,9% das disciplinas são voltadas para a formação geral do profissional (17 disciplinas);

- d) 8,5% das disciplinas são voltadas para a área de Informática (6 disciplinas), e
- e) 4,2% das disciplinas são relacionadas a clientes (3 disciplinas).

Nesta divisão, as disciplinas foram alocadas segundo seu objetivo maior, não tendo sido levada em consideração a interdisciplinaridade existente entre os conteúdos, o que evidencia-se principalmente nas disciplinas práticas.

O profissional cursa estas disciplinas de acordo com a capacidade de oferta do Departamento de Ciência e Gestão da Informação, mas pode, ainda, direcionar-se para outras áreas, cursando disciplinas eletivas em qualquer outro departamento da Universidade Federal do Paraná.

A próxima seção, Pesquisa de Campo, evidencia que algumas destas habilidades são requisitos básicos para que este profissional esteja habilitado a trabalhar em ambiente de *data warehouse*, integrando-se a equipes multidisciplinares.

## 6 PESQUISA DE CAMPO

A necessidade de informações que complementassem a literatura pertinente sobre a questão dos profissionais envolvidos no projeto, implantação e acompanhamento de *data warehouses* levou à execução desta pesquisa de campo.

De acordo com os dados obtidos na primeira questão da entrevista, a composição das equipes mostrou uma variação significativa, pois em um dos casos há atualmente, apenas um analista responsável pelo projeto, enquanto nos demais há equipes estruturadas com mais funcionários, de acordo com o escopo do projeto. Foram citados os seguintes profissionais/funções envolvidos nos projetos: analista de sistemas, líder de equipe, analista de programas, gerente de desenvolvimento, *logistic* (que prepara a estrutura para acolher o *data warehouse*), *user education* (integra o usuário ao ambiente) e desenvolvedor de aplicações. Algumas destas denominações ocorrem em virtude da utilização da metodologia de desenvolvimento de projetos adotada por uma indústria de *softwares*.

Quanto à necessidade de incorporação de mais profissionais à equipe, há indicação de que isso pode ocorrer para suprir necessidades momentâneas, para melhor estruturação da equipe que está desenvolvendo o projeto e para ampliação destes projetos com a incorporação de outros projetos afins. Dentro das necessidades citadas estão: analistas de sistemas, administradores de bancos de dados, usuários finais e às vezes, programadores.

Dentre as necessidades de conhecimento ou dificuldades da equipe estão a necessidade de acompanhamento constante das tecnologias que são desenvolvidas para o ambiente de *data warehouse* e da área de tecnologias de bancos de dados, ferramentas de metadados, ferramentas de análise e levantamento das reais necessidades dos usuários finais ou analistas de negócios para a estruturação do *data warehouse*.

As competências e habilidades apresentadas como necessárias de serem incorporadas às equipes são: poder de comunicação, capacidade de extração das reais

necessidades dos usuários, conhecimentos de bancos de dados relacionais e modelagem dimensional de dados e metodologia de planejamento, acompanhamento e gerência de projetos. Foi detectada também a necessidade de pessoas com conhecimento do negócio da empresa, ou capazes de analisar profundamente o negócio, e pessoas com conhecimento para a análise de metadados, proporcionando ao usuário final, além da informação, o seu significado e sua conceituação e a solução completa ao seu problema de informação.

Quanto à qualidade dos dados em termos de adoção de padrões, não foram citadas normas internacionais, como as estabelecidas pela Organização Internacional de Padronização (ISO) para tecnologias da informação, ou normas da Associação Brasileira de Normas Técnicas (ABNT) para desenvolvimento e qualidade de *software*, ou ambas para a área de Documentação.

Verificou-se, portanto, que a padronização dos dados é trabalhada, mas de acordo com critérios estabelecidos internamente em cada organização. Em dois dos casos entrevistados estes critérios são normatizados, em função da preocupação com a qualidade dos processos de desenvolvimento de projetos de bancos de dados e com a adoção de um padrão institucional. Nestes casos, todo trabalho de consistência dos dados, (limpeza e tratamento) é realizado com base nestes padrões internos.

No terceiro caso, o entrevistado afirmou que os dados são normalizados, mas não normatizados, ficando esta tarefa sob a responsabilidade e *feeling* do analista para resolver problemas de inconsistências.

## 7 CONSIDERAÇÕES FINAIS

As exposições realizadas neste relatório procuraram fornecer subsídios para a identificação do espaço de atuação do Gestor da Informação em bancos de dados do tipo *data warehouse*.

Este tipo de banco de dados é uma evolução das ferramentas de suporte à tomada de decisão.

Entende-se por sistema de apoio à decisão, aquele que auxilia os executivos na tomada de decisões, provendo-os com dados referentes ao desempenho da organização, permitindo combinações desses elementos através da utilização das tecnologias da informação.

O *data warehouse* é uma das ferramentas utilizadas nesse processo e pode ser definido como um grande banco de dados composto pelo histórico de atuação da organização e que serve como base para traçar suas estratégias de atuação.

Com tantas informações disponíveis nestes bancos de dados, o desafio que se coloca para as organizações atualmente, não é o do acesso à informação, mas de decifrar seu valor estratégico, para usá-la no momento adequado.

Assim como surgem constantemente novas tecnologias, surgem também profissionais com novos perfis que atendam as necessidades por elas criadas, tornando as áreas do conhecimento e o campo para sua atuação cada vez mais especializados.

O Gestor da Informação é resultante dessa necessidade de profissionais demandada pelas organizações. Este profissional pode ser definido como o responsável pela gerência e fornecimento de produtos e serviços de informação em uma organização, quer seja no nível tático, operacional ou gerencial, em consonância com as necessidades de seus clientes.

Diante do perfil profissional do Gestor da Informação e das necessidades observadas em pesquisa de campo, percebe-se que há espaço de atuação para este profissional em ambientes de *data warehouse*.

As colocações deste profissional podem se dar sob dois aspectos:

- a) atuação voltada para o cliente, e
- b) atuação voltada à informação.

Na atuação voltada para o cliente, o Gestor da Informação pode ter como atribuições o levantamento das reais necessidades dos usuários e avaliação da interface usuário-computador. Quando se realiza este levantamento, obtém-se um perfil das necessidades de acesso às informações, o que fornece subsídios para o estabelecimento dos níveis de segurança do banco de dados, evitando problemas com espionagem e roubo de informações. Estas colocações são corroboradas pela definição de Gestor da Informação apresentada por FONTES (2001, p. 2) e citada na seção 5, em que o Gestor teria atribuições de guarda da informação em uma organização.

O Gestor da Informação mostra-se apto a desempenhar estas atividades pois, três disciplinas (4,2%) estão relacionadas a clientes, sendo que duas delas são obrigatórias.

Na atuação voltada à informação especificamente, o Gestor pode ter como atribuições a participação na elaboração do projeto do *data warehouse*, na estruturação dos dados e de sua documentação e no processo de padronização dos dados antes de sua carga no *data warehouse*, além da normatização dos procedimentos.

O Gestor da Informação tem uma formação privilegiada quanto ao aspecto de elaboração de projetos, pois além das três disciplinas obrigatórias que o capacitam a realizar seu trabalho de conclusão de curso por meio de elaboração e aplicação de projeto, há outras disciplinas que também instruem e exigem a elaboração de projetos em seus trabalhos curriculares.

Da mesma forma, a estruturação e padronização de dados e procedimentos são abordadas em duas disciplinas específicas que tratam da representação da informação e da utilização de padrões, Análise e Representação Descritiva da Informação e Normalização Documentária, sendo estes conteúdos requisitados aos alunos no decorrer do Curso, em disciplinas posteriores.

O tema banco de dados, indispensável para o trabalho com *data warehouse*, é tratado em disciplina específica, Laboratório de Bancos de Dados, além de estar embutido diretamente em disciplinas voltadas ao tratamento da informação e



indiretamente em outras cujo instrumental seja bancos de dados, totalizando seis (8,4% das disciplinas), entendendo, pois, que os alunos recebem uma bagagem de conhecimentos da área, suficiente para sua introdução em equipes interdisciplinares voltadas à aplicação de tecnologias da informação.

Fica evidenciado, portanto, que há espaço para o Gestor da Informação, cabendo a este procurar qualificar-se continuamente para atuar no ambiente de *data warehouse*, no entanto, cabe ainda neste trabalho a apresentação de algumas recomendações direcionadas ao Curso de Gestão da Informação, ao Gestor da Informação e à continuidade de estudos sobre o tema.

Visando criar uma política de introdução dos graduandos nos conceitos de sistemas de apoio à decisão e as ferramentas nele utilizadas, recomenda-se a realização de seminários, a serem proferidos por profissionais que atuem na área de inteligência competitiva.

A introdução de disciplinas que abordem técnicas de entrevista e cursos de extensão que tenham por objetivo aperfeiçoar o poder de comunicação dos profissionais, também mostra-se fundamental, uma vez que a pesquisa de campo demonstrou esta carência.

Modelagem relacional *versus* multidimensional de bancos de dados e o estudo de metadados também devem ser introduzidos nas disciplinas de Informática, se a intenção é capacitar o profissional para a atuação em *data warehouse* ou sistemas de apoio à tomada de decisão.

A apresentação de ferramentas que possam ser utilizadas na elaboração, acompanhamento e documentação de projetos como o *MSProject*, entre outros, mostram-se indispensáveis para que os Gestores da Informação estejam mais capacitados a gerenciar projetos.

Aos Gestores da Informação recomenda-se o aprofundamento nos conceitos de inteligência competitiva e ferramentas de suporte à tomada de decisões, por meio de cursos de especialização em Informática, Inteligência Competitiva e Gestão do

Conhecimento, além da participação em seminários e grupos de discussão, para que o profissional esteja completamente integrado ao tema.

Dentre os temas que podem oferecer continuidade à pesquisa, destacam-se:

- a) estudo de outras ferramentas utilizadas no suporte à tomada de decisões;
- b) levantamento de termos utilizados em ambiente de *data warehouse* para elaboração de glossários de termos técnicos;
- c) levantamento dos instrumentos normativos, nacionais e internacionais, que possam ser aplicados em ambiente de *data warehouse*;
- d) estudo comparativo entre duas organizações que tenham implantado o *data warehouse* e avaliação dos resultados obtidos;
- e) estudo teórico-conceitual sobre inteligência competitiva (*business intelligence*).

O tema *data warehouse* é amplo e complexo, exigindo a realização de trabalho interdisciplinar para que as organizações alcancem os objetivos estabelecidos em seus projetos, quando decidem pela implantação de uma ferramenta como esta.

Evidencia-se, portanto, que a utilização das tecnologias de *software* não resolve por si a problemática informacional do corpo gerencial das organizações. Agregada a estas tecnologias, necessariamente devem estar presentes a análise semântica, econômica, estatística, sociológica, entre outras conveniências, para que a apresentação ou disponibilização dos dados ultrapasse a mera compressão, granularidade, integridade ou demais recursos de tratamento de dados. A Implementação de recursos inteligentes para a **análise automática dos dados** exige, com certeza, a participação humana para a interpretação e determinação prévia das evidências que devem ser destacadas entre os dados, ressaltando pontos críticos e merecedores de atenção, o que mostra, mais uma vez, a necessidade de pessoas com conhecimentos do negócio da empresa, habilidades técnicas especializadas e capacidade de percepção superior aos conhecimentos fragmentados por cada profissão.

## REFERÊNCIAS

ANDREATTO, Ricardo. **Construindo um data warehouse e analisando suas informações com data mining e OLAP**. Valinhos (SP), 1999. 46f. Monografia (Graduação em Ciência da Computação) – Faculdade de Ciências Administrativas de Valinhos. Disponível em <<http://www.datawarehouses.hpg.ig.com.br/>> Acesso em: 07 jan. 2002.

BERNARDINO, Fábio Pereira. **Gerência de metadados: como integrar seu data warehouse**. Rio de Janeiro, 1998. 17f. Monografia – Universidade Federal do Rio de Janeiro.

BISPO, Carlos Alberto Ferreira. Data warehouse. In: \_\_\_\_\_. **Uma análise da nova geração de sistemas de apoio à decisão**. São Carlos (SP), 1998. 23f. Dissertação (Mestrado em Engenharia da Produção) – Escola de Engenharia de São Carlos, Universidade de São Paulo. Disponível em: <<http://cazarini.cpd.eesc.sc.usp.br/Bispo/DI/Capitulo3.html>> Acesso em: 08 jan. 2002.

CANONGIA, Claudia *et al.* Convergência da inteligência competitiva com construção de visão de futuro: proposta metodológica de Sistema de Informação Estratégica (SIE). **DataGramaZero – Revista de Ciência da Informação**, v. 2, n. 3, jun. 2001. 15p. Disponível em: <[http://www.dgz.org.br/jun01/Art\\_02.htm](http://www.dgz.org.br/jun01/Art_02.htm)> Acesso em: 19 dez. 2001.

CARVALHO, Kátia. Disseminação de informação e informação de inteligência organizacional. **DataGramaZero – Revista de Ciência da Informação**, v. 2, n. 3, jun. 2001. Disponível em: <[http://www.dgz.org.br/jun01/Art\\_04.htm](http://www.dgz.org.br/jun01/Art_04.htm)> Acesso em: 19 dez. 2001.

CAUTELA, Alciney Lourenço; POLLONI, Enrico Giulio Franco. **Sistemas de informação na administração de empresas**. São Paulo: Atlas, 1996. 244p.

FONTES, Edson. **Segurança da informação: investimento ou custo operacional?** Disponível em: <<http://www.securenet.com.br/artigo.php?artigo=108>> Acesso em: 09 jan. 2002.

FURLAN, José Davi. **Reengenharia da informação: do mito à realidade**. São Paulo: Makron Books, 1994. 132p.

GIL, Antonio Carlos. Entrevista. In: \_\_\_\_\_. **Métodos e técnicas de pesquisa social**. 5. ed. São Paulo: Atlas, 1999. p. 117-127.

GUIA do acadêmico de Gestão da Informação. Curitiba: UFPR, 1998. 32p.

HISTÓRICO do data warehouse. 3p. Disponível em: <[http://www.dwbrasil.com.br/html/artdw\\_hist.html](http://www.dwbrasil.com.br/html/artdw_hist.html)> Acesso em: 14 dez. 2001.

INMON, William H. **Como construir o data warehouse**. Rio de Janeiro: Campus, 1997. 388p.

INMON, William H.; TERDEMAN, R. H.; IMHOFF, Claudia. **Data warehousing: como transformar informações em oportunidades de negócios**. São Paulo: Berkeley, 2001. 266p.

JAMIL, George Leal.; NEVES, Jorge Tadeu de Ramos. A era da informação: considerações sobre o desenvolvimento de tecnologias da informação. **Perspect. Cienc. Inf.**, Belo Horizonte, v. 5, n. 1, p. 41-53, jan./jun. 2000.

LE COADIC, Yves François. **A Ciência da Informação**. Brasília: Briquet de Lemos, 1996. 116p.

MIORELLI, Henrique Salatino. Proposta de um roteiro para projetar um data warehouse. **Bate Byte**, Curitiba, n. 100, ago. 2000. Disponível em:  
<<http://www.pr.gov.br/celepar/celepar/batebyte/edicoes/2000/bb100/proposta.htm>> Acesso em: 28 mai. 2001.

MUSICANTE, Martin A. (ed). **Data warehousing e OLAP**. Curitiba: UFPR, 2001. 161p.

NORTON, Peter. **Introdução à informática**. São Paulo: Atlas, 1996. 244p.

OLIVEIRA, Djalma de Pinho Rebouças. **Sistemas de informações gerenciais**: estratégicas, táticas, operacionais. 7. ed. rev. atual. São Paulo: Atlas, 2001. 285p.

PEREIRA, Maria José Lara de Bretas; FONSECA, João Gabriel Marques. **Faces da decisão**: as mudanças de paradigmas e o poder da decisão. São Paulo: Makron Books, 1997. 275p.

POLÍTICA de segurança da informação. 4p. Disponível em:  
<[http://br.geocities.com/jasonbs\\_1917/seguranca/politica.htm](http://br.geocities.com/jasonbs_1917/seguranca/politica.htm)> Acesso em: 09 jan. 2001.

POLLONI, Enrico Giulio Franco. **Administrando sistemas de informação**. 2. ed. São Paulo: Futura, 2000. 272 p.

REZENDE, Denis Alcides; ABREU, Aline França de. **Tecnologia da informação aplicada a sistemas de informações empresariais**: o papel da informação e dos sistemas de informações nas empresas. São Paulo: Atlas, 2000. 311p.

SANTOS, Aldemar Araújo. **Informática na empresa**. São Paulo: Atlas, 1998. 619p.

WG SYSTEMS. **Dicionário de BI**. Disponível em:  
<[http://www.wgsystems.com.br/bi/bi\\_dicionario.htm](http://www.wgsystems.com.br/bi/bi_dicionario.htm)> Acesso em: 07 jan. 2001.

YAMADA, Cibele Keiko; GIOVANELI, Fabíola Wolf. **Data warehouse**. Curitiba, 1997. 63f. Monografia (Especialização em Redes e Sistemas Distribuídos) – Pontifícia Universidade Católica do Paraná.

## DOCUMENTOS CONSULTADOS

- AMOR, Daniel. **A (r)evolução do e-business**: vivendo e trabalhando em um mundo interconectado. São Paulo: Makron Books, 2000.
- APPLEHANS, Wayne; GLOBE, Alden; LAUGERO, Greg. **Managing knowledge**: a practical web-based approach. Boston: Addison-Wesley, 1999. 115p.
- CHENG, Paul S.; CHANG, Pintsang. Transforming corporate information into value through data warehousing and data mining. **Aslib Proceedings**, v. 50, n. 5, p. 109-113, may 1998.
- COMPUTERWORLD. Suplemento especial data warehouse. Rio de Janeiro: IDG, 1998.
- DATA mining. Disponível em: <<http://www.dwbrasil.com.br/html/dmining.html>> Acesso em: 14 dez. 2001.
- DAVENPORT, Thomas H. **Ecologia da informação**: por que só a tecnologia não basta para o sucesso na Era da Informação. 3. ed. São Paulo: Futura, 2000. 316p.
- INMON, William H.; WELCH, J. D.; GLASSEY, Katherine L. **Gerenciando data warehouse**. São Paulo: Makron Books, 1999. 375p.
- KALAKOTA, Ravi; ROBINSON, Marcia. **E-business**: roadmap for successs. 6. ed. Massachussets: Longman, 1999.
- KIMBALL, Ralph. **The data warehouse toolkit**: practical techniques for building dimensional Data Warehouses. New York: J. Wiley, 374p.
- LAUDON, Kenneth C.; LAUDON, Janet P. **Management information systems**: organization and technology. 3rd. ed. New Jersey: Prentice-Hall, 1994. 818p.
- LODI, João Bosco. **A entrevista**: teoria e prática. 7. ed. São Paulo: Pioneira, 1991. 176p.
- MARCO, David. **Top 10 mistakes to avoid when implementing a meta data repository**. Disponível em: <<http://www.datawarehouse.com/iknowledge/articles/print/cfm?ContentID=1291>> Acesso em: 18 abr. 2001.
- MCGEE, James V.; PRUSSAK, Lawrence. **Gerenciamento estratégico da informação**: aumente a competitividade e a eficiência de sua empresa utilizando a informação como uma ferramenta estratégica. 6. ed. Rio de Janeiro: Campus, 1994. 244p.
- METADADOS. Disponível em: <<http://www.dwbrasil.com.br/html/metadata.html>> Acesso em: 14 dez. 2001.
- PEARSON, David. Marketing for survival. **CIO Magazine**, 15 Apr. 1998. Disponível em: <[http://www.cio.com/archive/041598\\_finger.html](http://www.cio.com/archive/041598_finger.html)> Acesso em: 4 fev. 2001.
- PONJUÁN DANTE, Glória. **Gestión de información en las organizaciones**: principios, conceptos y aplicaciones. Santiago de Chile: CECAPI, 1998. 222p.
- REBOUÇAS, Lidia. 3 aulas digitais. **Negócios Exame**, São Paulo, p. 26-37, nov. 2001.
- ROBERTS-WITT, Sarah L. Gold diggers. **PC Magazine**, London, v. 20, n. 4, p. 5-10, Feb. 2001.

SAPIRO, Arão. Inteligência empresarial: a revolução informacional da ação competitiva. **Rev. de Adm. de Empresas**, São Paulo, v. 33, n. 3, p. 106-124, maio/jun. 1993.

SOCIEDADE da informação no Brasil: livro verde. Brasília: Ministério da Ciência e Tecnologia, 2000. 195p.

TARAPANOFF, Kira; ARAUJO JUNIOR, Rogério Henrique; CORNIER, Patrícia Marie Jeanne. Sociedade da informação e inteligência em unidades de informação. **Ci. Inf.**, Brasília, v. 29, n. 3, p. 91-100, set./dez. 2000.

## **APÊNDICE – FORMULÁRIO DA ENTREVISTA**

UNIVERSIDADE FEDERAL DO PARANÁ  
Setor de Ciências Sociais Aplicadas  
Departamento de Ciência e Gestão da Informação  
Curso de Gestão da Informação  
Disciplina: Pesquisa em Informação II  
Orientadora: Profª. Ligia Leindorf Bartz Kraemer  
Aluno: Adriano Lopes  
Pesquisa sobre *Data Warehouse*

### **Formulário da entrevista**

#### **Identificação**

Nome completo  
Curso de graduação  
Pós-graduação  
Empresa  
Cargo ocupado atualmente

#### **Questões específicas**

- 1) Qual a composição da equipe responsável pela implantação e pelo acompanhamento do *data warehouse*?  
(*função e quantidade*)
- 2) Há necessidade de incorporação de mais profissionais? Quantos e quais as funções?
- 3) Quais as carências de conhecimento e/ou dificuldades da equipe?
- 4) Quais as competências e/ou habilidades a serem incorporadas à equipe?
- 5) Foram ou são utilizados padrões na implantação e na manutenção do data warehouse? Quais?
- 6) Como é trabalhada a qualidade dos dados em termos de padronização?



**ANEXO – CURRÍCULO PLENO DO CURSO DE GESTÃO DA  
INFORMAÇÃO – RESOLUÇÃO N. 24/98 - CEPE**